

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2003, 10:54:49 ; Search time 52.4483 Seconds  
(without alignments)  
99.084 Million cell updates/sec

Title: US-09-003-869-3

Perfect score: 120

Sequence: 1 XXGTXXXKXQEEEAVALRLLXXXXLNGKXSSGAXXXX 39

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

all number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | ID             | Description        |
|------------|-------|-------------|----------------|--------------------|
| 1          | 94    | 78.3        | 37 20 AAY17618 | Exendin agonist pe |
| 2          | 94    | 78.3        | 37 22 AAE08527 | Exendin agonist pe |
| 3          | 94    | 78.3        | 37 22 AAB64353 | Exendin agonist, S |
| 4          | 94    | 78.3        | 39 21 AAB11313 | Exendin agonist pe |
| 5          | 94    | 78.3        | 39 22 AAE08383 | Exendin agonist pe |
| 6          | 93    | 77.5        | 36 20 AAY17606 | Exendin agonist pe |
| 7          | 93    | 77.5        | 36 21 AAB11263 | Exendin agonist pe |
| 8          | 93    | 77.5        | 36 21 AAB53029 | Amino acid sequenc |
| 9          | 93    | 77.5        | 36 21 AAY94184 | Exendin agonist pe |
| 10         | 93    | 77.5        | 36 22 AAE08515 | Exendin agonist pe |

|    |    |      |                |                    |
|----|----|------|----------------|--------------------|
| 11 | 93 | 77.5 | 36 22 AAB64351 | Exendin agonist, S |
| 12 | 93 | 77.5 | 37 20 AAY24869 | Exendin agonist pe |
| 13 | 93 | 77.5 | 37 20 AAY24853 | Exendin agonist pe |
| 14 | 93 | 77.5 | 37 20 AAY24854 | Exendin agonist pe |
| 15 | 93 | 77.5 | 37 21 AAB11275 | Exendin agonist pe |
| 16 | 93 | 77.5 | 37 21 AAB53041 | Exendin agonist c  |
| 17 | 93 | 77.5 | 37 21 AAY94196 | Amino acid sequenc |
| 18 | 93 | 77.5 | 37 22 AAE08427 | Exendin agonist pe |
| 19 | 93 | 77.5 | 37 22 AAE08428 | Exendin agonist pe |
| 20 | 93 | 77.5 | 37 22 AAE08443 | Exendin agonist pe |
| 21 | 93 | 77.5 | 37 22 AAB64263 | Exendin agonist, S |
| 22 | 93 | 77.5 | 37 22 AAB64264 | Exendin agonist, S |
| 23 | 93 | 77.5 | 37 22 AAB64279 | Exendin agonist, S |
| 24 | 93 | 77.5 | 39 21 AAB11311 | Exendin agonist pe |
| 25 | 93 | 77.5 | 39 21 AAY94039 | Amino acid sequenc |
| 26 | 93 | 77.5 | 39 21 AAY94040 | Amino acid sequenc |
| 27 | 93 | 77.5 | 39 21 AAY94043 | Amino acid sequenc |
| 28 | 93 | 77.5 | 39 22 AAE08379 | Exendin agonist pe |
| 29 | 93 | 77.5 | 39 22 AAE08380 | Exendin agonist pe |
| 30 | 93 | 77.5 | 39 22 AAE08381 | Exendin agonist, S |
| 31 | 93 | 77.5 | 39 22 AAB64219 | Exendin agonist pe |
| 32 | 92 | 76.7 | 35 20 AAY31535 | Exendin agonist pe |
| 33 | 92 | 76.7 | 35 20 AAY24839 | Exendin agonist pe |
| 34 | 92 | 76.7 | 35 21 AAB11161 | Exendin agonist pe |
| 35 | 92 | 76.7 | 35 21 AAB11265 | Exendin agonist pe |
| 36 | 92 | 76.7 | 35 21 AAB53020 | Exendin agonist c  |
| 37 | 92 | 76.7 | 35 21 AAB53031 | Exendin agonist c  |
| 38 | 92 | 76.7 | 35 21 AAY94074 | Amino acid sequenc |
| 39 | 92 | 76.7 | 35 21 AAY94186 | Amino acid sequenc |
| 40 | 92 | 76.7 | 35 21 AAE08413 | Exendin agonist pe |
| 41 | 92 | 76.7 | 35 22 AAE08517 | Exendin agonist pe |
| 42 | 92 | 76.7 | 35 22 AAB64249 | Exendin agonist, S |
| 43 | 92 | 76.7 | 35 22 AAB64353 | Exendin agonist, S |
| 44 | 92 | 76.7 | 35 22 AAB64353 | Exendin agonist, S |
| 45 | 92 | 76.7 | 36 20 AAY31533 | Exendin agonist pe |

#### ALIGNMENTS

RESULT 1  
AAY17618  
ID AAY17618 standard; peptide; 37 AA.  
XX  
AC AAY17618;  
XX  
DT 09-AUG-1999 (first entry)  
XX  
DE Exendin agonist peptide #84.  
XX  
DE Exendin; agonist; Heloderma sp.; Gila monster; venom; lizard;  
KW diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;  
KW hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.  
XX  
OS Synthetic.  
OS Heloderma sp.  
XX  
PN WO9925728-A1.  
XX  
PD 27-MAY-1999.  
XX  
PF 13-NOV-1998; 98WO-US24273.  
XX  
PR 14-NOV-1997; 97US-0066029.  
PA (AMYL-) AMYLIN PHARM INC.  
XX  
PI Beeley NRA, Prickett KS;  
XX  
DR WPI; 1999-347456/29.  
XX  
PT Peptide agonists of exendin - delay stomach emptying, for treating diabetes and hypo- or hyper-glycaemia

XX  
XS  
PS  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
SO

Claim 28; Fig 4; 144pp; English.

AAV17535 to AAV17624 represent extendin peptide agonists. Extendins are CC peptides that are found in the venom of the Gila-monster, a lizard CC endogenous to Arizona and Northern Mexico. The peptide agonists are CC used to treat diabetes mellitus (types I or II), hyperglycaemia or CC hypoglycaemia. They can also be used for in vitro and in vivo studies CC on extendins and their agonists. They regulate gastric motility and slow CC gastric emptying (resulting in lower post-prandial glucose levels).

Sequence 37 AA;

Query Match 78.3%; Score 94; DB 20; Length 37;  
Best Local Similarity 65.6%; Pred. No. 4.4e-10;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 4 GTXXXXXSKQEEAEVRLXXXLXNGGXSGA 35  
|| ||| ||||||| ||||| ||||  
4 GTFTSALKSQMEEAERVLFLEWLNKGASGSA 35

RESULT 2  
AAE08527  
ID AAE08527 standard; peptide; 37 AA.  
XX AC AAE08527;  
DT 01-NOV-2001 (first entry)  
XX DE Extendin agonist peptide #172.  
XX DE  
XX KW Extendin agonist; antilipemic; cardiast; triglyceride; inotropic;  
KW diuretic; coronary heart disease; dyslipidaemia.  
XX OS Synthetic.  
FH Key Location/Qualifiers  
FI Modified-site 31 /note= "N-methyl alanine"  
FT Modified-site 36 /note= "N-methyl alanine"  
FT Modified-site 37 /note= "N-methyl alanine"  
FT FT /note= "N-methyl alanine; C-terminal amide"  
PN WO200151078-AL.  
XX 19-JUL-2001.  
  
09-JAN-2001; 2001WO-US00719.  
10-JAN-2000; 2000US-0175365.  
(AMYL-) AMYLIN PHARM INC.  
Kolterman OG, Young AA;  
WPI; 2001-514422/56.  
  
Use of extendin and extendin agonist compounds for modulating  
triglyceride levels, and treating heart disease and dyslipidemia -  
Example 178; Page 143; 161pp; English.

The patent discloses a method for modulating plasma or postprandial  
triglyceride and other lipid levels by administering extendin or an  
extendin agonist. Extendins have inotropic and diuretic effects. They  
suppress the secretion of glucagon. Extendin and its agonists have  
a significant effect on the reduction of blood serum triglyceride  
concentrations. They are used to treat coronary heart disease and  
dyslipidaemia, and for modifying postprandial triglyceride levels.  
The present peptide sequence is an agonist of extendin.

|          |  |                                     |
|----------|--|-------------------------------------|
| SQL      | Sequence   | 37 AA;                              |
|          | Query Match  | 78.3%; Score 94; DB 22; Length 37;  |
|          | Best Local Similarity  | 65.6%; Pred: No. 4.4e-10;           |
|          | Matches 21; Conservative   | 0; Mismatches 11; Indels 0; Gaps 0; |
| QY       | 4 GTYXXXXKQEEAEAVRLXXIXLKNGXSSGA   | 35                                  |
|          | (( (( (((((((((( (((((( ((((((((   |                                     |
| DB       | 4 GTFTSALKSQMEEEAEVRLFTEWLKNKGASSGA  | 35                                  |
|          | (( (( (((((((((( (((((( ((((((((   |                                     |
| RESULT 3 |  |                                     |
| AAB64363 | ID AAB64363 standard; peptide; 37 AA.                                      |                                     |
| XX       | AAB64363;  |                                     |
| XX       | AC   |                                     |
| XX       | DT   |                                     |
| XX       | 27-MAR-2001 (first entry)  |                                     |
| XX       | Exendin agonist, SEQ ID NO:183.  |                                     |
| DE       | XX   |                                     |
| XX       | Exendin agonist: gestational diabetes mellitus; GDM; insulin resistance;   |                                     |
| KW       | pregnancy complication; neonatal abnormality; blood glucose modulator;     |                                     |
| KW       | insulinotropic; anorectic; exendin-4.                                      |                                     |
| KW       | XX   |                                     |
| OS       | Holoderma suspectum.   |                                     |
| OS       | Synthetic.   |                                     |
| PN       | WO200073331-A2.  |                                     |
| PD       | 07-DEC-2000.   |                                     |
| XX       | XX   |                                     |
| Pf       | 23-MAY-2000: 2000WO-US14231.   |                                     |
| XX       | XX   |                                     |
| PR       | 01-JUN-1999; 99US-0323867.   |                                     |
| PA       | (AMYL-) AMYLIN PHARM INC.  |                                     |
| PI       | Hiles R, Prickett KS;  |                                     |
| DR       | WPI: 2001-137634/14.   |                                     |
| PT       | Use of extendins or extendin agonists for lowering or reducing blood       |                                     |
| PT       | glucose levels and treating gestational diabetes mellitus in a subject,    |                                     |
| PT       | especially in a human -  |                                     |
| PS       | Example 178; Page 119; 133pp; English.                                     |                                     |
| CC       | The invention relates to the use of an extendin (AAB64181-B64182) or       |                                     |
| CC       | an extendin agonist (AAB64185-B64368) for treating gestational diabetes    |                                     |
| CC       | mellitus (GDM) in a patient. GDM arises during pregnancy, and is due       |                                     |
| CC       | to a combination of increased insulin resistance and a diminished          |                                     |
| CC       | ability to increase insulin secretion. In contrast to a normal             |                                     |
| CC       | pregnancy, both insulin resistance and insulin secretion increase. GDM     |                                     |
| CC       | pregnancies are associated with complications in both the mother and the   |                                     |
| CC       | fetus. Women with GDM have increased rates of Caesarian delivery,          |                                     |
| CC       | hypertensive disorders such as pre-eclampsia, and urinary tract            |                                     |
| CC       | infections. GDM results in an elevated rate of foetal abnormalities such   |                                     |
| CC       | as neural tube defects, and is associated with an increased risk of        |                                     |
| CC       | neonatal morbidities such as hypoglycaemia, hypocalcaemia,                 |                                     |
| CC       | hypomagnesaemia, polycythaemia, hyperbilirubinaemia, and subsequent        |                                     |
| CC       | childhood and adolescent obesity. Extendins are peptides from the salivary |                                     |
| CC       | secretions of the gila monster (exendin-4) and the Mexican beaded lizard   |                                     |
| CC       | (exendin-3) which exhibit homology with several members of the             |                                     |
| CC       | glucagon-like peptide family, particularly GLP-1, and have similar         |                                     |
| CC       | insulinotropic effects. Unlike the compounds used to treat type 2          |                                     |
| CC       | diabetes, which are contraindicated for GDM, extendins and extendin        |                                     |
| CC       | agonists do not cross the placenta and thus do not cause severe prolonged  |                                     |
| CC       | hypoglycaemia in the newborn. They have a potent and prolonged effect on   |                                     |
| CC       | blood glucose, and, unlike conventional insulin therapy, should not cause  |                                     |
| CC       | weight gain, as they inhibit gastric emptying and reduce appetite. The     |                                     |
| CC       | present sequence represents a extendin agonist of the invention which is   |                                     |
| CC       | based upon the sequence of exendin-4.                                      |                                     |

XX Sequence 37 AA;  
 Query Match 78.3%; Score 94; DB 22; Length 37;  
 Best Local Similarity 65.6%; Pred. No. 4.7e-10;  
 Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 4 GTXXXXXKQEEAEVRLXXXXLKNKGXSSGA 35  
 II III IIIIIII IIIII IIII  
 DB 4 GTFTSLSKQEEAEVRLFIEFLKNGGASSGA 35

RESULT 4  
 AAE08383  
 ID AAE08383 standard; Peptide: 39 AA.  
 XX  
 AC AAE08383;  
 XX  
 DT 20-FEB-2001 (first entry)  
 extendin agonist peptide SEQ ID NO 39.  
 KW Extendin; agonist; treatment; antidiabetic; insulin sensitivity; diabetes;  
 KW plasma glucose; gastric emptying; food intake.  
 XX Synthetic.  
 XX WO200041546-A2.  
 PN 20-JUL-2000.  
 PD  
 XX 10-JAN-2000; 2000US-0116380.  
 PF 14-JAN-1999; 99US-0116380.  
 PR (AMYL-) AMYLIN PHARM INC.  
 PA Young A, L'italien JJ, Kolterman O;  
 PI WPI; 2000-514584/46.  
 DR  
 XX New formulations comprising an extendin or extendin agonist peptide used  
 PT for increasing the sensitivity of a subject to insulin to treat  
 PT diabetes -  
 XX  
 PS Example 44; Figure 15; 281pp; English.  
 CC This invention describes a novel formulation (I) comprising an extendin or  
 CC extendin agonist peptide, a buffer and an iso-osmolality modifier which  
 CC has a pH of 3-7. The products of the invention have antidiabetic  
 CC activity. The extendin or extendin agonist is used to increase the  
 CC sensitivity of a subject to insulin to treat diabetes and disorders which  
 CC would benefit from agents which lower plasma glucose levels and disorders  
 CC which would benefit from agents that delay and/or slow gastric emptying  
 CC or reducing food intake.  
 XX Sequence 39 AA;  
 Query Match 78.3%; Score 94; DB 21; Length 39;  
 Best Local Similarity 65.6%; Pred. No. 4.7e-10;  
 Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 4 GTXXXXXKQEEAEVRLXXXXLKNKGXSSGA 35  
 II III IIIIIII IIIII IIII  
 DB 4 GTFTSLSKQEEAEVRLFIEFLKNGGASSGA 35

RESULT 5  
 AAE08383  
 ID AAE08383 standard; peptide: 39 AA.  
 XX  
 AC AAE08383;  
 XX  
 KW Extendin; agonist; Heloderma sp.; Gila monster; venom; lizard;

DT 01-NOV-2001 (first entry)  
 XX Extendin agonist peptide #30.  
 DE  
 XX  
 KW Extendin agonist; antilipemic; cardiant; triglyceride; inotropic;  
 KW diuretic; coronary heart disease; dyslipidaemia.  
 XX Synthetic.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 31 /note= "N-Methyl-alanine"  
 FT Modified-site 36 /note= "N-Methyl-alanine"  
 FT Modified-site 37 /note= "N-Methyl-alanine"  
 FT Modified-site 38 /note= "N-Methyl-alanine"  
 FT Modified-site 39 /note= "N-Methyl-alanine"  
 FT Modified-site 39 /note= "C-terminal amide"  
 XX WO200151078-A1.  
 PN 19-JUL-2001.  
 PD  
 XX 09-JAN-2001; 2001WO-US00719.  
 PF 10-JAN-2000; 2000US-0175365.  
 PR (AMYL-) AMYLIN PHARM INC.  
 PA Kolterman OG, Young AA;  
 PI WPI; 2001-514422/56.  
 DR  
 XX Use of extendin and extendin agonist compounds for modulating  
 PT triglyceride levels, and treating heart disease and dyslipidemia -  
 PT Example 30; Page -: 161pp; English.  
 PS  
 CC The patent discloses a method for modulating plasma or postprandial  
 CC triglyceride and other lipid levels by administering extendin or an  
 CC extendin agonist. Extendins have inotropic and diuretic effects. They  
 CC suppress the secretion of glucagon. Extendin and its agonists have  
 CC a significant effect on the reduction of blood serum triglyceride  
 CC concentrations. They are used to treat coronary heart disease and  
 CC dyslipidaemia, and for modifying postprandial triglyceride levels.  
 CC The present peptide sequence is an agonist of extendin.  
 CC Note: The present sequence is not shown in the specification but is  
 CC derived from SEQ ID NO:3 shown in page 17 of the specification.  
 XX Sequence 39 AA;  
 Query Match 78.3%; Score 94; DB 22; Length 39;  
 Best Local Similarity 65.6%; Pred. No. 4.7e-10;  
 Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 4 GTXXXXXKQEEAEVRLXXXXLKNKGXSSGA 35  
 II III IIIIIII IIIII IIII  
 DB 4 GTFTSLSKQEEAEVRLFIEFLKNGGASSGA 35

RESULT 6  
 AAY17606  
 ID AAY17606 standard; peptide: 36 AA.  
 XX  
 AC AAY17606;  
 XX  
 DT 09-AUG-1999 (first entry)  
 XX Extendin agonist peptide #72.  
 DE  
 XX  
 KW Extendin; agonist; Heloderma sp.; Gila monster; venom; lizard;

KW diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;  
 KW hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.

XX Synthetic.  
 OS Heloderma sp.

XX WO9925728-A1.

XX 27-MAY-1999.

XX 13-NOV-1998; 98WO-US24273.

XX 14-NOV-1997; 97US-0066029.

XX (AMYL-) AMYLIN PHARM INC.

XX Beeley NRA, Prickett KS;

XX WPI; 1999-347456/29.

PT Peptide agonists of exendin - delay stomach emptying, for treating  
 XX diabetes and hypo- or hyper-glycaemia

PS Claim 28; Fig 4; 144pp; English.

XX AAY17535 to AAY17624 represent exendin peptide agonists. Exendins are  
 CC peptides that are found in the venom of the Gila-monster, a lizard  
 CC endogenous to Arizona and Northern Mexico. The peptide agonists are  
 CC used to treat diabetes mellitus (types I or II), hyperglycaemia or  
 CC hypoglycaemia. They can also be used for in vitro and in vivo studies  
 CC on exendins and their agonists. They regulate gastric motility and slow  
 CC gastric emptying (resulting in lower post-prandial glucose levels).

XX Sequence 36 AA;

Query Match 77.5%; Score 93; DB 20; Length 36;

Best Local Similarity 65.6%; Pred. No. 6.5e-10;

Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXKQEEAEVRLXXXXXKNGXSSGA 35

DB 4 GTFTSDASKQLEEEAVRLFIEFLKNGPSSGA 35

RESULT 7

AAB11263

ID AAB11263 standard; Peptide; 36 AA.

AAB11263;

DT 20-FEB-2001 (first entry)

DE exendin agonist peptide SEQ ID NO 171.

XX Exendin; agonist; treatment; antidiabetic; insulin sensitivity; diabetes;  
 KW plasma glucose; gastric emptying; food intake.

XX Synthetic.

XX WO200041546-A2.

XX 20-JUL-2000.

XX 10-JAN-2000; 2000US-0116380.

XX 14-JAN-1999; 99US-0116380.

XX (AMYL-) AMYLIN PHARM INC.

XX Young A, L'Italien JJ, Kolterman O;

XX WPI; 2000-514584/46.

PT New formulations comprising an exendin or exendin agonist peptide used  
 PT for increasing the sensitivity of a subject to insulin to treat  
 PT diabetes -

XX Example 180; Page 229; 281pp; English.

XX This invention describes a novel formulation (I) comprising an exendin or  
 CC exendin agonist peptide, a buffer and an iso-osmolality modifier which  
 CC has a pH of 3-7. The products of the invention have antidiabetic  
 CC activity. The exendin or exendin agonist is used to increase the  
 CC sensitivity of a subject to insulin to treat diabetes and disorders which  
 CC would benefit from agents which lower plasma glucose levels and disorders  
 CC which would benefit from agents that delay and/or slow gastric emptying  
 CC or reducing food intake.

XX Sequence 36 AA;

Query Match 77.5%; Score 93; DB 21; Length 36;

Best Local Similarity 65.6%; Pred. No. 6.5e-10;

Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXKQEEAEVRLXXXXXKNGXSSGA 35

DB 4 GTFTSDASKQLEEEAVRLFIEFLKNGPSSGA 35

RESULT 8

AAB53029

ID AAB53029 standard; Peptide; 36 AA.

XX AAB53029;

XX 28-FEB-2001 (first entry)

XX Extendin agonist compound #157.

XX Extendin; agonist; diabetes; obesity; eating disorder;  
 KW dyslipidaemia; insulin-resistance syndrome; food intake.

XX Heloderma sp.

XX WO200066629-A1.

XX 09-NOV-2000.

XX 28-APR-2000; 2000WO-US11814.

XX 30-APR-1999; 99US-0132018.

XX (AMYL-) AMYLIN PHARM INC.

XX Young A, Prickett K;

XX WPI; 2000-672834/65.

XX Modified exendin or an exendin agonist linked to one or more  
 PT polyethylene glycol (PEG) polymers, modulate plasma glucose levels,  
 PT useful for treating disorders such as diabetes and obesity -

XX Disclosure; Fig 4; 119pp; English.

XX The present invention relates to extendins and their agonists which have  
 CC been modified with molecular weight increasing agents such as  
 CC polyethylene glycol (PEG). These can be used in the treatment of  
 CC diabetes, obesity, impaired glucose tolerance, postprandial dumping  
 CC syndrome, postprandial hyperglycaemia, eating disorders, insulin  
 CC resistance syndrome, dyslipidaemia and to suppress glucagon secretion.

XX Sequence 36 AA;

Query Match 77.5%; Score 93; DB 21; Length 36;

Best Local Similarity 65.6%; Pred. No. 6.5e-10;

Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;



OY 4 GTXXXXXKQXEEAVRLXXXXLKNKGXSSGA 35  
 || ||| ||||| ||||| ||||| |||||  
 DD 4 GTFTSDASKQLEEEAVRLFTEFLKNGPSSGA 35

## RESULT 9

AAAY94184  
 ID AAAY94184 standard; peptide; 36 AA.

AC AAAY94184;  
 DT 20-OCT-2000 (first entry)  
 DE Amino acid sequence of an extendin agonist.

XX Extendin; Gila monster lizard; Mexican Beaded lizard; agonist;  
 KW glucagon-like peptide; plasma glucagon; necrolytic erythema;  
 KW glucagonoma; hyperglucagonemia; diabetes.

Synthetic.  
 Heloderma sp.

XX Key Location/Qualifiers  
 FH Modified-site 36  
 FT /note= "amidated residue"

XX WO200041548-A2.

XX 20-JUL-2000.

XX 14-JAN-2000; 2000WO-US00942.

XX 14-JAN-1999; 99US-0116380.

PR 30-APR-1999; 99US-0132017.

PR 10-JAN-2000; 2000US-0175365.

XX (AMYL-) AMYLIN PHARM INC.

XX Young A, Gedulin B;

XX WPI; 2000-490999/43.

XX Lowering plasma glucagon using extendin, an extendin agonist, a modified  
 PT extendin or a modified extendin agonist, useful for treating  
 PT hyperglucagonemia and diabetes -

XX Disclosure; Fig 4G; 96pp; English.

PS The present sequence represents a modified extendin or extendin agonist.  
 CC Extendins are found in the salivary glands of the Gila monster and  
 CC Mexican Beaded lizard, and have sequence similarity to glucagon-like  
 CC peptides. They are used in the method of the invention. The specification  
 CC describes a method for lowering plasma glucagon, comprising administering  
 CC an extendin, an extendin agonist, a modified extendin or a modified extendin  
 CC agonist. These compounds lower plasma glucagon level. The method is  
 CC useful for lowering plasma glucagon in subjects, preferably humans,  
 CC suffering from necrolytic erythema or glucagonoma. The method is also  
 CC useful for treating hyperglucagonemia and other conditions that would  
 CC benefit from reduced glucagon levels and/or suppression of glucagon,  
 CC e.g. type 1 and type 2 diabetes.

XX Sequence 36 AA;

Query Match 77.5%; Score 93; DB 21; Length 36;  
 Best Local Similarity 65.6%; Pred. No. 6.5e-10;  
 Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 4 GTXXXXXKQXEEAVRLXXXXLKNKGXSSGA 35  
 || ||| ||||| ||||| ||||| |||||

DD 4 GTFTSDASKQLEEEAVRLFTEFLKNGPSSGA 35

## RESULT 10

AAE08515  
 ID AAE08515 standard; peptide; 36 AA.

XX AAE08515;

XX 01-NOV-2001 (first entry)

XX Extendin agonist peptide #160.

XX Extendin agonist; antilipemic; cardiant; triglyceride; inotropic;  
 KW diuretic; coronary heart disease; dyslipidaemia.

XX Synthetic.

XX Key Location/Qualifiers  
 FH Modified-site 36  
 FT /note= "C-terminal amide"

XX WO200151078-A1.

XX 19-JUL-2001.

XX 09-JAN-2001; 2001WO-US00719.

PR 10-JAN-2000; 2000US-0175365.

XX (AMYL-) AMYLIN PHARM INC.

XX Kolterman OG, Young AA;

XX WPI; 2001-514422/56.

XX Use of extendin and extendin agonist compounds for modulating  
 PT triglyceride levels, and treating heart disease and dyslipidemia -  
 XX Example 166; Page 136; 161pp; English.

XX The patent discloses a method for modulating plasma or postprandial  
 CC triglyceride and other lipid levels by administering extendin or an  
 CC extendin agonist. Extendins have inotropic and diuretic effects. They  
 CC suppress the secretion of glucagon. Extendin and its agonists have  
 CC a significant effect on the reduction of blood serum triglyceride  
 CC concentrations. They are used to treat coronary heart disease and  
 CC dyslipidaemia, and for modifying postprandial triglyceride levels.  
 CC The present peptide sequence is an agonist of extendin.

XX Sequence 36 AA;

Query Match 77.5%; Score 93; DB 22; Length 36;  
 Best Local Similarity 65.6%; Pred. No. 6.5e-10;  
 Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 4 GTXXXXXKQXEEAVRLXXXXLKNKGXSSGA 35  
 || ||| ||||| ||||| ||||| |||||

DB 4 GTFTSDASKQLEEEAVRLFTEFLKNGPSSGA 35

## RESULT 11

AAB64351  
 ID AAB64351 standard; peptide; 36 AA.

XX AAB64351;

XX 27-MAR-2001 (first entry)

XX Extendin agonist, SEQ ID NO:171.

XX Extendin agonist; gestational diabetes mellitus; GDM; insulin resistance;  
 KW pregnancy complication; neonatal abnormality; blood glucose modulator;  
 KW insulinotropic; anorectic; extendin-4.

XX Heloderma suspectum.

OS Synthetic.  
 XX WO200073331-A2.  
 XX  
 XX PD 07-DEC-2000.  
 XX  
 XX 23-MAY-2000; 2000WO-US14231.  
 XX PF  
 XX 01-JUN-1999; 99US-0323867.  
 XX PR  
 XX (AMYL-) AMYLIN PHARM INC.  
 XX PA  
 XX Hiles R, Prickett KS;  
 XX PI  
 XX WPI; 2001-137634/14.  
 XX DR  
 XX  
 XX Use of extendin or extendin agonists for lowering or reducing blood  
 PT glucose levels and treating gestational diabetes mellitus in a subject,  
 PT especially in a human -  
 XX  
 XX Example 166; Page 113; 133pp; English.  
 XX  
 XX The invention relates to the use of an extendin (AAB64181-B64182) or  
 CC an extendin agonist (AAB64185-B64368) for treating gestational diabetes  
 CC mellitus (GDM) in a patient. GDM arises during pregnancy, and is due  
 CC to a combination of increased insulin resistance and a diminished  
 CC ability to increase insulin secretion. In contrast, in a normal  
 CC pregnancy, both insulin resistance and insulin secretion increase. GDM  
 CC pregnancies are associated with complications in both the mother and the  
 CC foetus. Women with GDM have increased rates of Caesarian delivery,  
 CC hypertensive disorders such as pre-eclampsia, and urinary tract  
 CC infections. GDM results in an elevated rate of foetal abnormalities such  
 CC as neural tube defects, and is associated with an increased risk of  
 CC neonatal morbidities such as hypoglycaemia, hypocalcaemia, subsequent  
 CC hypomagnesaemia, polycythaemia, hyperbilirubinaemia, and subsequent  
 CC childhood and adolescent obesity. Extendins are peptides from the salivary  
 CC secretions of the Gila monster (extendin-4) and the Mexican bearded lizard  
 CC (extendin-3) which exhibit homology with several members of the  
 CC glucagon-like peptide family, particularly GLP-1, and have similar  
 CC insulinotropic effects. Unlike the compounds used to treat type 2  
 CC diabetes, which are contraindicated for GDM, extendins and extendin  
 CC agonists do not cross the placenta and thus do not cause severe prolonged  
 CC hypoglycaemia in the newborn. They have a potent and prolonged effect on  
 CC blood glucose, and, unlike conventional insulin therapy, should not cause  
 CC weight gain, as they inhibit gastric emptying and reduce appetite. The  
 CC present sequence represents a extendin agonist of the invention which is  
 CC based upon the sequence of extendin-4.  
 CC  
 CC Sequence 36 AA;  
 CC  
 CC Query Match 77.5%; Score 93; DB 22; Length 36;  
 CC Best Local Similarity 65.6%; Pred. No. 6.5e-10;  
 CC Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
 CC  
 CC QY 4 GTXXXXXKQEEAEVRLXXXXLKNKGXSSGA 35  
 CC || ||| ||||| ||||| |||||  
 CC Db 4 GTTSDASKOLEEAEVRLFIETLKNKGXSSGA 35  
 CC  
 CC RESULT 12  
 CC AAY24869  
 CC ID AAY24869 standard; peptide; 37 AA.  
 CC XX  
 CC AC AAY24869;  
 CC XX  
 CC DT 24-AUG-1999 (first entry)  
 CC XX  
 CC DE Extendin agonist peptide #61.  
 CC  
 CC Extendin; agonist; Heloderma sp.; Gila monster; venom; lizard;  
 CC diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;  
 CC hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.  
 CC  
 CC OS Synthetic.  
 CC OS Heloderma sp.  
 CC PN WO9925727-A2.  
 CC XX  
 CC PD 27-MAY-1999.  
 CC XX  
 CC PF 13-NOV-1998; 98WO-US24210.  
 CC XX  
 CC PR 14-NOV-1997; 97US-0065442.  
 CC XX  
 CC PA (AMYL-) AMYLIN PHARM INC.  
 CC XX  
 CC PI Hiles R, Prickett KS;  
 CC XX  
 CC DR WPI; 1999-394773/33.  
 CC XX  
 CC New extendin agonist peptides - can regulate gastric motility and  
 PT slow gastric emptying, used for treating, e.g. diabetes  
 PT  
 XX Claim 18; Fig 4; 108pp; English.  
 XX  
 XX AAY24809 to AAY24877 represent extendin agonist peptides which can  
 CC regulate gastric motility and slow gastric emptying. The peptides can be  
 CC used for treating e.g. diabetes or hyperglycaemic or hypoglycaemic  
 CC conditions. The peptides are extendin agonists which have activity as  
 CC agents to regulate gastric motility and to slow gastric emptying, as  
 CC evidenced by the ability to reduce post-prandial glucose levels in  
 CC mammals. They can be used for the treatment of Type I and II diabetes and  
 CC hyperglycaemic or hypoglycaemic conditions. They can also be used for the  
 CC treatment of disorders which would be benefited by agents which lower  
 CC plasma glucose levels and in treatment of disorders which would be  
 CC benefited with agents useful in delaying and/or slowing gastric  
 CC emptying.  
 CC  
 CC XX Sequence 37 AA;  
 CC  
 CC Query Match 77.5%; Score 93; DB 20; Length 37;  
 CC Best Local Similarity 68.8%; Pred. No. 6.7e-10;  
 CC Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
 CC  
 CC QY 4 GTXXXXXKQEEAEVRLXXXXLKNKGXSSGA 35  
 CC || ||| ||||| ||||| |||||  
 CC Db 4 GTTSDASKOLEEAEVRLFIETLKNKGXSSGA 35  
 CC  
 CC RESULT 13  
 CC AAY24853  
 CC ID AAY24853 standard; peptide; 37 AA.  
 CC XX  
 CC AC AAY24853;  
 CC XX  
 CC DT 24-AUG-1999 (first entry)  
 CC XX  
 CC DE Extendin agonist peptide #45.  
 CC  
 CC Extendin; agonist; Heloderma sp.; Gila monster; venom; lizard;  
 CC diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;  
 CC hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.  
 CC  
 CC OS Synthetic.  
 CC OS Heloderma sp.  
 CC PN WO9925727-A2.  
 CC XX  
 CC PD 27-MAY-1999.  
 CC XX  
 CC PF 13-NOV-1998; 98WO-US24210.  
 CC XX  
 CC PR 14-NOV-1997; 97US-0065442.  
 CC XX  
 CC PA (AMYL-) AMYLIN PHARM INC.  
 CC XX

PI Beeley NRA, Prickett KS;  
 XX WPI; 1999-394773/33.  
 XX  
 PT New extendin agonist peptides - can regulate gastric motility and  
 PT slow gastric emptying, used for treating, e.g. diabetes  
 XX  
 PS Claim 18; Fig 4; 108pp; English.  
 XX  
 CC AAY24809 to AAY24877 represent extendin agonist peptides which can  
 CC regulate gastric motility and slow gastric emptying. The peptides can be  
 CC used for treating e.g. diabetes or hyperglycaemic or hypoglycaemic  
 CC conditions. The peptides are extendin agonists which have activity as  
 CC agents to regulate gastric motility and to slow gastric emptying, as  
 CC evidenced by the ability to reduce post-prandial glucose levels in  
 CC mammals. They can be used for the treatment of Type I and II diabetes and  
 CC hyperglycaemic or hypoglycaemic conditions. They can also be used for the  
 CC treatment of disorders which would be benefited by agents which lower  
 CC plasma glucose levels and in treatment of disorders which would be  
 CC benefited with agents useful in delaying and/or slowing gastric  
 CC emptying.  
 SQ Sequence 37 AA;  
 Query Match 77.5%; Score 93; DB 20; Length 37;  
 Best Local Similarity 65.6%; Pred. No. 6.7e-10;  
 Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
 QY 4 GTXXXXXKQEEAEVRLXXXXLKNKGXSSGA 35  
 II III IIIIIII IIIII IIII  
 Db 4 GTFTSLSKQMEEEAVRLFIEWLKNGGASSGA 35  
 RESULT 14  
 AAY24854  
 ID AAY24854 standard; peptide; 37 AA.  
 XX  
 AC AAY24854;  
 XX  
 DT 24-AUG-1999 (first entry)  
 XX  
 DE Extendin agonist peptide #46.  
 XX  
 KW Extendin; agonist: Heloderma sp.; Gila monster; venom; lizard;  
 KW diabetes mellitus type I; diabetes mellitus type II; hyperglycaemia;  
 KW hypoglycaemia; plasma glucose; gastric emptying; stomach emptying.  
 XX  
 OS Synthetic.  
 Heloderma sp.  
 WO9925727-A2.  
 XX  
 PD 27-MAY-1999.  
 XX  
 PF 13-NOV-1998; 98WO-US24210.  
 XX  
 PR 14-NOV-1997; 97US-0065442.  
 XX  
 PA (AMYL-) AMYLIN PHARM INC.  
 XX  
 PI Beeley NRA, Prickett KS;  
 XX  
 DR WPI; 1999-394773/33.  
 XX  
 PT New extendin agonist peptides - can regulate gastric motility and  
 PT slow gastric emptying, used for treating, e.g. diabetes  
 XX  
 PS Claim 18; Fig 4; 108pp; English.  
 XX  
 CC AAY24809 to AAY24877 represent extendin agonist peptides which can  
 CC regulate gastric motility and slow gastric emptying. The peptides can be  
 CC used for treating e.g. diabetes or hyperglycaemic or hypoglycaemic  
 CC conditions. The peptides are extendin agonists which have activity as

CC agents to regulate gastric motility and to slow gastric emptying, as  
 CC evidenced by the ability to reduce post-prandial glucose levels in  
 CC mammals. They can be used for the treatment of Type I and II diabetes and  
 CC hyperglycaemic or hypoglycaemic conditions. They can also be used for the  
 CC treatment of disorders which would be benefited by agents which lower  
 CC plasma glucose levels and in treatment of disorders which would be  
 CC benefited with agents useful in delaying and/or slowing gastric  
 CC emptying.  
 XX  
 SQ Sequence 37 AA;  
 Query Match 77.5%; Score 93; DB 20; Length 37;  
 Best Local Similarity 65.6%; Pred. No. 6.7e-10;  
 Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
 QY 4 GTXXXXXKQEEAEVRLXXXXLKNKGXSSGA 35  
 II III IIIIIII IIIII IIII  
 Db 4 GTFTSLSKQMEEEAVRLFIEWLKNGGASSGA 35  
 RESULT 15  
 AABL1275  
 ID AABL1275 standard; peptide; 37 AA.  
 XX  
 AC AABL1275;  
 XX  
 DT 20-FEB-2001 (first entry)  
 XX  
 DE extendin agonist peptide SEQ ID NO 183.  
 XX  
 KW Extendin; agonist: treatment; antidiabetic; insulin sensitivity; diabetes;  
 KW plasma glucose; gastric emptying; food intake.  
 XX  
 OS Synthetic.  
 WO200041546-A2.  
 XX  
 PN 20-JUL-2000.  
 XX  
 PD 10-JAN-2000; 2000US-0116380.  
 XX  
 PF 14-JAN-1999; 99US-0116380.  
 XX  
 PR (AMYL-) AMYLIN PHARM INC.  
 XX  
 PA Young A, L'Italien JJ, Kolterman O;  
 PI WPI; 2000-514584/46.  
 XX  
 DR New formulations comprising an extendin or extendin agonist peptide used  
 XX for increasing the sensitivity of a subject to insulin to treat  
 XX diabetes -  
 PT  
 XX Example 192; Page 238; 281pp; English.  
 XX  
 CC This invention describes a novel formulation (I) comprising an extendin or  
 CC extendin agonist peptide, a buffer and an iso-osmolality modifier which  
 CC has a pH of 3-7. The products of the invention have antidiabetic  
 CC activity. The extendin or extendin agonist is used to increase the  
 CC sensitivity of a subject to insulin to treat diabetes and disorders which  
 CC would benefit from agents which lower plasma glucose levels and disorders  
 CC which would benefit from agents that delay and/or slow gastric emptying  
 CC or reducing food intake.  
 XX  
 SQ Sequence 37 AA;  
 Query Match 77.5%; Score 93; DB 21; Length 37;  
 Best Local Similarity 68.8%; Pred. No. 6.7e-10;  
 Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
 QY 4 GTXXXXXKQEEAEVRLXXXXLKNKGXSSGA 35  
 II III IIIIIII IIIII IIII  
 Db 4 GTFTSLSKQMEEEAVRLFIEWLKNGGASSGA 35

Wed Jan 8 11:04:34 2003

us-09-003-869-3.rag

Page 8

Search completed: January 8, 2003, 10:58:21  
Job time : 53.4483 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2003, 10:54:49 ; Search time 11.0948 Seconds  
(without alignments)  
103.426 Million cell updates/sec

Title: US-09-003-869-3  
Perfect score: 120  
Sequence: 1 XXXGTXKXQXEEAEVRLXXXXLKNKGXSGAXXXX 39

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 2942292 residues  
11 number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID                  | Description       |
|------------|-------|-------------|--------|---------------------|-------------------|
| 1          | 91    | 75.8        | 39     | 1 US-08-066-480-1   | Sequence 1, Appl  |
| 2          | 91    | 75.8        | 39     | 1 US-08-066-480-2   | Sequence 2, Appl  |
| 3          | 91    | 75.8        | 39     | 4 US-09-302-596-7   | Sequence 7, Appl  |
| 4          | 91    | 75.8        | 39     | 4 US-09-302-596-9   | Sequence 9, Appl  |
| 5          | 91    | 75.8        | 39     | 4 US-09-623-618B-11 | Sequence 11, Appl |
| 6          | 91    | 75.8        | 39     | 4 US-09-623-618B-12 | Sequence 12, Appl |
| 7          | 91    | 75.8        | 39     | 4 US-09-333-415-7   | Sequence 7, Appl  |
| 8          | 91    | 75.8        | 39     | 4 US-09-333-415-9   | Sequence 9, Appl  |
| 9          | 91    | 75.8        | 39     | 4 US-09-303-016-7   | Sequence 7, Appl  |
| 10         | 91    | 75.8        | 39     | 4 US-09-303-016-9   | Sequence 9, Appl  |
| 11         | 91    | 75.8        | 40     | 4 US-09-623-618B-18 | Sequence 18, Appl |
| 12         | 91    | 75.8        | 40     | 4 US-09-623-618B-19 | Sequence 19, Appl |
| 13         | 91    | 75.8        | 40     | 4 US-09-623-618B-31 | Sequence 31, Appl |
| 14         | 91    | 75.8        | 40     | 4 US-09-623-618B-32 | Sequence 32, Appl |
| 15         | 91    | 75.8        | 40     | 4 US-09-623-618B-33 | Sequence 33, Appl |
| 16         | 91    | 75.8        | 40     | 4 US-09-623-618B-34 | Sequence 34, Appl |
| 17         | 83    | 69.2        | 31     | 1 US-08-066-480-5   | Sequence 5, Appl  |
| 18         | 83    | 69.2        | 31     | 4 US-09-302-596-8   | Sequence 8, Appl  |
| 19         | 83    | 69.2        | 31     | 4 US-09-623-618B-15 | Sequence 15, Appl |
| 20         | 83    | 69.2        | 31     | 4 US-09-623-618B-24 | Sequence 24, Appl |
| 21         | 83    | 69.2        | 31     | 4 US-09-333-415-8   | Sequence 8, Appl  |
| 22         | 83    | 69.2        | 31     | 4 US-09-303-016-8   | Sequence 8, Appl  |
| 23         | 75    | 62.5        | 31     | 1 US-08-066-480-3   | Sequence 3, Appl  |
| 24         | 75    | 62.5        | 31     | 1 US-08-066-480-4   | Sequence 4, Appl  |
| 25         | 75    | 62.5        | 31     | 4 US-09-623-618B-14 | Sequence 14, Appl |
| 26         | 75    | 62.5        | 31     | 4 US-09-623-618B-23 | Sequence 23, Appl |
| 27         | 75    | 62.5        | 32     | 4 US-09-623-618B-35 | Sequence 35, Appl |

|    |      |      |     |   |                   |                   |
|----|------|------|-----|---|-------------------|-------------------|
| 28 | 62.5 | 52.1 | 29  | 4 | US-09-623-618B-22 | Sequence 22, Appl |
| 29 | 59.5 | 49.6 | 31  | 4 | US-09-623-618B-13 | Sequence 13, Appl |
| 30 | 57.5 | 47.9 | 30  | 4 | US-09-623-618B-21 | Sequence 21, Appl |
| 31 | 57.5 | 47.9 | 31  | 4 | US-09-623-618B-20 | Sequence 20, Appl |
| 32 | 41   | 34.2 | 516 | 4 | US-09-154-750A-86 | Sequence 86, Appl |
| 33 | 41   | 34.2 | 589 | 2 | US-08-317-305-2   | Sequence 2, Appl  |
| 34 | 41   | 34.2 | 589 | 2 | US-08-317-305-4   | Sequence 4, Appl  |
| 35 | 41   | 34.2 | 589 | 3 | US-08-862-508-2   | Sequence 2, Appl  |
| 36 | 41   | 34.2 | 589 | 3 | US-08-862-508-4   | Sequence 4, Appl  |
| 37 | 41   | 34.2 | 589 | 5 | PCT-US95-12508-2  | Sequence 2, Appl  |
| 38 | 39   | 32.5 | 341 | 1 | US-08-062-024B-5  | Sequence 5, Appl  |
| 39 | 39   | 32.5 | 341 | 1 | US-08-891-254-5   | Sequence 5, Appl  |
| 40 | 39   | 32.5 | 341 | 2 | US-08-756-407-5   | Sequence 5, Appl  |
| 41 | 39   | 32.5 | 341 | 2 | US-08-819-539-5   | Sequence 5, Appl  |
| 42 | 39   | 32.5 | 341 | 2 | US-09-030-270A-5  | Sequence 5, Appl  |
| 43 | 39   | 32.5 | 341 | 4 | US-08-984-207-5   | Sequence 5, Appl  |
| 44 | 39   | 32.5 | 341 | 4 | US-09-013-587-5   | Sequence 5, Appl  |
| 45 | 39   | 32.5 | 341 | 4 | US-09-013-587-5   | Sequence 5, Appl  |

ALIGNMENTS

RESULT 1  
US-08-066-480-1  
; Sequence 1, Application US/08066480  
; Patent No. 5424286  
; GENERAL INFORMATION:  
; APPLICANT: Eng, John  
; TITLE OF INVENTION: Pharmaceutical Compositions And Use of  
; TITLE OF INVENTION: Exendin-3 and Exendin-4 for treatment of Diabetes Mellitus  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Allegretti & Witcoff, Ltd.  
; STREET: 10 S. Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/066,480  
; FILING DATE: 24-MAR-1993  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McDonnell, John J  
; REGISTRATION NUMBER: 26,949  
; REFERENCE/DOCKET NUMBER: 93,084  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 39 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; MOLECULE TYPE: peptide  
; NAME/KEY: peptide  
; LOCATION: 1..39  
; OTHER INFORMATION: /label= Exendin-3  
US-08-066-480-1

Query Match 75.8%; Score 91; DB 1; Length 39;  
Best Local Similarity 65.6%; Pred. No. 8.7e-10;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXKQXEEAEVRLXXXXLKNKGXSGA 35



; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: Peptide  
US-09-623-618B-11

Query Match 75.8%; Score 91; DB 4; Length 39;  
Best Local Similarity 65.6%; Pred. No. 8.7e-10;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
QY 4 GTYXXXXXKQXEEAVRLXXXXXKNGXSSGA 35  
|| ||| ||||| ||||| ||||| |||||  
DB 4 GTFTSDLSKQWEEAVRLFTWLKNGPSSGA 35

RESULT 6  
US-09-623-618B-12  
; Sequence 12, Application US/09623618B  
; Patent No. 6329336  
; GENERAL INFORMATION:  
; APPLICANT: Bridon, Dominique P.  
; APPLICANT: L'Archeveque, Benoit  
; APPLICANT: Ezrin, Alan M.  
; APPLICANT: Holmes, Darren L.  
; APPLICANT: Leblanc, Anouk  
; APPLICANT: St. Pierre, Serge  
; TITLE OF INVENTION: LONG LASTING INSULINOTROPIC PEPTIDES  
; FILE REFERENCE: 500862001620  
; CURRENT FILING DATE: 2000-09-05  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: PCT/US00/13563  
; PRIOR FILING DATE: 2000-05-17  
; PRIOR APPLICATION NUMBER: 60/159,783  
; PRIOR FILING DATE: 1999-10-15  
; PRIOR APPLICATION NUMBER: 60/134,406  
; PRIOR FILING DATE: 1999-05-17  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 39  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: Peptide  
US-09-623-618B-12

Query Match 75.8%; Score 91; DB 4; Length 39;  
Best Local Similarity 65.6%; Pred. No. 8.7e-10;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
QY 4 GTYXXXXXKQXEEAVRLXXXXXKNGXSSGA 35  
|| ||| ||||| ||||| ||||| |||||  
DB 4 GTFTSDLSKQWEEAVRLFTWLKNGPSSGA 35

RESULT 7  
US-09-333-415-7  
; Sequence 7, Application US/09333415  
; Patent No. 6344180  
; GENERAL INFORMATION:  
; APPLICANT: Holst, Jens J.  
; APPLICANT: Vilsboll, Tina  
; TITLE OF INVENTION: GLP-1 as a Diagnostic Test to Determine Beta-Cell  
; TITLE OF INVENTION: Function and the Presence of the Condition of IGT and  
; TITLE OF INVENTION: Type-II Diabetes  
; FILE REFERENCE: P03987US0  
; CURRENT APPLICATION NUMBER: US/09/333,415  
; CURRENT FILING DATE: 1999-06-15  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 39

; TYPE: PRT  
; ORGANISM: Heloderma suspectum  
US-09-333-415-7  
Query Match 75.8%; Score 91; DB 4; Length 39;  
Best Local Similarity 65.6%; Pred. No. 8.7e-10;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
QY 4 GTYXXXXXKQXEEAVRLXXXXXKNGXSSGA 35  
|| ||| ||||| ||||| ||||| |||||  
DB 4 GTFTSDLSKQWEEAVRLFTWLKNGPSSGA 35

RESULT 8  
US-09-333-415-9  
; Sequence 9, Application US/09333415  
; Patent No. 6344180  
; GENERAL INFORMATION:  
; APPLICANT: Holst, Jens J.  
; APPLICANT: Vilsboll, Tina  
; TITLE OF INVENTION: GLP-1 as a Diagnostic Test to Determine Beta-Cell  
; TITLE OF INVENTION: Function and the Presence of the Condition of IGT and  
; TITLE OF INVENTION: Type-II Diabetes  
; FILE REFERENCE: P03987US0  
; CURRENT APPLICATION NUMBER: US/09/333,415  
; CURRENT FILING DATE: 1999-06-15  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 39  
; TYPE: PRT  
; ORGANISM: Heloderma suspectum  
US-09-333-415-9

Query Match 75.8%; Score 91; DB 4; Length 39;  
Best Local Similarity 65.6%; Pred. No. 8.7e-10;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
QY 4 GTYXXXXXKQXEEAVRLXXXXXKNGXSSGA 35  
|| ||| ||||| ||||| ||||| |||||  
DB 4 GTFTSDLSKQWEEAVRLFTWLKNGPSSGA 35

RESULT 9  
US-09-303-016-7  
; Sequence 7, Application US/09303016  
; Patent No. 6429197  
; GENERAL INFORMATION:  
; APPLICANT: Coolidge, Thomas R.  
; APPLICANT: Ehlers, Mario R.W.  
; TITLE OF INVENTION: Metabolic Intervention with GLP-1 or its Biologically  
; TITLE OF INVENTION: Active Analogues to Improve the Function of the  
; TITLE OF INVENTION: Ischemic and Reperfused Brain  
; FILE REFERENCE: P03660US2  
; CURRENT APPLICATION NUMBER: US/09/303,016  
; CURRENT FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/103,498  
; PRIOR FILING DATE: 1998-10-08  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 39  
; TYPE: PRT  
; ORGANISM: Heloderma suspectum  
US-09-303-016-7

Query Match 75.8%; Score 91; DB 4; Length 39;  
Best Local Similarity 65.6%; Pred. No. 8.7e-10;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
QY 4 GTYXXXXXKQXEEAVRLXXXXXKNGXSSGA 35  
|| ||| ||||| ||||| ||||| |||||  
DB 4 GTFTSDLSKQWEEAVRLFTWLKNGPSSGA 35

```
RESULT 10
US-09-303-016-9
; Sequence 9, Application US/09303016
; Patent No. 6429197
; GENERAL INFORMATION:
; APPLICANT: Coolidge, Thomas R.
; APPLICANT: Ehlers, Mario R.W.
; TITLE OF INVENTION: Metabolic Intervention with GLP-1 or its Biologically
; TITLE OF INVENTION: Active Analogues to Improve the Function of the
; TITLE OF INVENTION: Ischemic and Reperfused Brain
; FILE REFERENCE: P03660US2
; CURRENT APPLICATION NUMBER: US/09/303,016
; CURRENT FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/103,498
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; EQ ID NO 9
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Heloderma suspectum
US-09-303-016-9

Query Match 75.8%; Score 91; DB 4; Length 39;
Best Local Similarity 65.6%; Pred. No. 8.7e-10;
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXKQXEEAVRLXXXXXKLNKGXSSGA 35
|| ||| ||||| ||||| |||||
Db 4 GTFTDLSKQMEEEAVRLFIEWLKNGPSSGA 35

RESULT 11
US-09-623-618B-18
; Sequence 18, Application US/09623618B
; Patent No. 6329336
; GENERAL INFORMATION:
; APPLICANT: Bridon, Dominique P.
; APPLICANT: L'Archeveque, Benoit
; APPLICANT: Ezrin, Alan M.
; APPLICANT: Holmes, Darren L.
; APPLICANT: Leblanc, Anouk
; APPLICANT: St. Pierre, Serge
; TITLE OF INVENTION: LONG LASTING INSULINOTROPIC PEPTIDES
; FILE REFERENCE: 500862001620
; CURRENT APPLICATION NUMBER: US/09/623,618B
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: PCT/US00/13563
; PRIOR FILING DATE: 2000-05-17
; PRIOR FILING DATE: 1999-10-15
; PRIOR FILING DATE: 1999-05-17
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSEQ for Windows Version 4.0
; EQ ID NO 18
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-09-623-618B-18

Query Match 75.8%; Score 91; DB 4; Length 40;
Best Local Similarity 65.6%; Pred. No. 8.9e-10;
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXKQXEEAVRLXXXXXKLNKGXSSGA 35
|| ||| ||||| ||||| |||||
Db 4 GTFTDLSKQMEEEAVRLFIEWLKNGPSSGA 35

RESULT 12
US-09-623-618B-19
; Sequence 19, Application US/09623618B
; Patent No. 6329336
; GENERAL INFORMATION:
; APPLICANT: Bridon, Dominique P.
; APPLICANT: L'Archeveque, Benoit
; APPLICANT: Ezrin, Alan M.
; APPLICANT: Holmes, Darren L.
; APPLICANT: Leblanc, Anouk
; APPLICANT: St. Pierre, Serge
; TITLE OF INVENTION: LONG LASTING INSULINOTROPIC PEPTIDES
; FILE REFERENCE: 500862001620
; CURRENT APPLICATION NUMBER: US/09/623,618B
; CURRENT FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: PCT/US00/13563
; PRIOR FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: 60/159,783
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: 60/134,406
; PRIOR FILING DATE: 1999-05-17
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSEQ for Windows Version 4.0
; EQ ID NO 19
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-09-623-618B-19

Query Match 75.8%; Score 91; DB 4; Length 40;
Best Local Similarity 65.6%; Pred. No. 8.9e-10;
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXKQXEEAVRLXXXXXKLNKGXSSGA 35
|| ||| ||||| ||||| |||||
Db 4 GTFTDLSKQMEEEAVRLFIEWLKNGPSSGA 35

RESULT 13
US-09-623-618B-31
; Sequence 31, Application US/09623618B
; Patent No. 6329336
; GENERAL INFORMATION:
; APPLICANT: Bridon, Dominique P.
; APPLICANT: L'Archeveque, Benoit
; APPLICANT: Ezrin, Alan M.
; APPLICANT: Holmes, Darren L.
; APPLICANT: Leblanc, Anouk
; APPLICANT: St. Pierre, Serge
; TITLE OF INVENTION: LONG LASTING INSULINOTROPIC PEPTIDES
; FILE REFERENCE: 500862001620
; CURRENT APPLICATION NUMBER: US/09/623,618B
; CURRENT FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: PCT/US00/13563
; PRIOR FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: 60/159,783
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: 60/134,406
; PRIOR FILING DATE: 1999-05-17
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSEQ for Windows Version 4.0
; EQ ID NO 31
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-09-623-618B-31

Query Match 75.8%; Score 91; DB 4; Length 40;
Best Local Similarity 65.6%; Pred. No. 8.9e-10;
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXKQXEEAVRLXXXXXKLNKGXSSGA 35
|| ||| ||||| ||||| |||||
Db 4 GTFTDLSKQMEEEAVRLFIEWLKNGPSSGA 35
```











```
; Sequence 36, Application US/09003869A
; Patent No. US20020137666A1
; GENERAL INFORMATION:
; APPLICANT: BEELEY, NIGEL ROBERT ARNOLD
; APPLICANT: PRICKETT, KATHRYN S.
; TITLE OF INVENTION: USE OF EXTENDINS AND AGONISTS THEREOF FOR
; TITLE OF INVENTION: THE REDUCTION OF FOOD INTAKE
; FILE REFERENCE: 231/181
; CURRENT APPLICATION NUMBER: US/09/003,869A
; EARLIER FILING DATE: 1998-01-07
; EARLIER APPLICATION NUMBER: US 60/034,905
; EARLIER FILING DATE: 1997-01-07
; EARLIER APPLICATION NUMBER: US 60/055,404
; EARLIER FILING DATE: 1997-08-08
; EARLIER APPLICATION NUMBER: US 60/065,442
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: US 60/066,029
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: artificially synthesized sequence of novel extendin agonist
; OTHER INFORMATION: compound
; FEATURE:
; OTHER INFORMATION: Xaa in positions 31, 36, 37 and 38 stands for homoproline.
; NAME/KEY: AMIDATION
; LOCATION: (39)...(39)
; OTHER INFORMATION: amidated Ser (Serinamide)
; US-09-003-869-36

Query Match      77.5%; Score 93; DB 10; Length 39;
Best Local Similarity 68.8%; Pred. No. 4.7e-10;
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTXXXXXKQXEEAVRLXXXXXKNGXSSGA 35
   || ||| ||||| ||||| |||
Db 4 GTFTDLSKQLEBEAVRLFTEFLKNGXSSGA 35

RESULT 6
US-09-003-869-39
; Sequence 39, Application US/09003869A
; Patent No. US20020137666A1
; GENERAL INFORMATION:
; APPLICANT: BEELEY, NIGEL ROBERT ARNOLD
; APPLICANT: PRICKETT, KATHRYN S.
; TITLE OF INVENTION: USE OF EXTENDINS AND AGONISTS THEREOF FOR
; TITLE OF INVENTION: THE REDUCTION OF FOOD INTAKE
; FILE REFERENCE: 231/181
; CURRENT APPLICATION NUMBER: US/09/003,869A
; EARLIER FILING DATE: 1998-01-07
; EARLIER APPLICATION NUMBER: US 60/034,905
; EARLIER FILING DATE: 1997-01-07
; EARLIER APPLICATION NUMBER: US 60/055,404
; EARLIER FILING DATE: 1997-08-08
; EARLIER APPLICATION NUMBER: US 60/065,442
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: US 60/066,029
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: artificially synthesized sequence of novel extendin agonist
```

```
; OTHER INFORMATION: artificially synthesized sequence of novel extendin agonist
; OTHER INFORMATION: compound
; FEATURE:
; OTHER INFORMATION: Xaa in positions 31, 36, 37 and 38 stands for n-methylalanine
; FEATURE:
; NAME/KEY: AMIDATION
; LOCATION: (39)...(39)
; OTHER INFORMATION: amidated Ser (Serinamide)
; US-09-003-869-39

Query Match      77.5%; Score 93; DB 10; Length 39;
Best Local Similarity 68.8%; Pred. No. 4.7e-10;
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTXXXXXKQXEEAVRLXXXXXKNGXSSGA 35
   || ||| ||||| ||||| |||
Db 4 GTFTDLSKQLEBEAVRLFTEFLKNGXSSGA 35

RESULT 7
US-09-003-869-69
; Sequence 69, Application US/09003869A
; Patent No. US20020137666A1
; GENERAL INFORMATION:
; APPLICANT: BEELEY, NIGEL ROBERT ARNOLD
; APPLICANT: PRICKETT, KATHRYN S.
; TITLE OF INVENTION: USE OF EXTENDINS AND AGONISTS THEREOF FOR
; TITLE OF INVENTION: THE REDUCTION OF FOOD INTAKE
; FILE REFERENCE: 231/181
; CURRENT APPLICATION NUMBER: US/09/003,869A
; EARLIER FILING DATE: 1998-01-07
; EARLIER APPLICATION NUMBER: US 60/034,905
; EARLIER FILING DATE: 1997-01-07
; EARLIER APPLICATION NUMBER: US 60/055,404
; EARLIER FILING DATE: 1997-08-08
; EARLIER APPLICATION NUMBER: US 60/065,442
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: US 60/066,029
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 69
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: artificially synthesized sequence of novel extendin agonist
; OTHER INFORMATION: compound
; FEATURE:
; NAME/KEY: AMIDATION
; LOCATION: (35)...(35)
; OTHER INFORMATION: amidated Ala (Alaninamide)
; US-09-003-869-69

Query Match      76.7%; Score 92; DB 10; Length 35;
Best Local Similarity 65.6%; Pred. No. 6.2e-10;
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXKQXEEAVRLXXXXXKNGXSSGA 35
   || ||| ||||| ||||| |||
Db 4 GTFTDLSKQLEBEAVRLFTEFLKNGXSSGA 35

RESULT 8
US-09-003-869-173
; Sequence 173, Application US/09003869A
; Patent No. US20020137666A1
; GENERAL INFORMATION:
; APPLICANT: BEELEY, NIGEL ROBERT ARNOLD
; APPLICANT: PRICKETT, KATHRYN S.
; TITLE OF INVENTION: USE OF EXTENDINS AND AGONISTS THEREOF FOR
```



```
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: US 60/066,029
; EARLIER FILING DATE: 1997-11-14
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 170
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: artificially synthesized sequence of novel extendin agonist
; OTHER INFORMATION: compound
; FEATURE:
; LOCATION: (36)...(36)
; OTHER INFORMATION: amidated Pro (Prolinamide)
US-09-003-869-170

Query Match          76.7%; Score 92; DB 10; Length 36;
Best Local Similarity 65.8%; Pred. No. 6.4e-10;
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTYXXXXXKQXEEAVRLXXXXXKLNKGXSSGA 35
   || ||| ||||| ||||| |||||
Db 4 GTFTSLSKQMEEEAVRLFTLEWLNKGXSSGA 35

RESULT 12
US-09-003-869-184
; Sequence 184, Application US/09003869A
; Patent No. US20020137666A1
; GENERAL INFORMATION:
; APPLICANT: BEELEY, NIGEL ROBERT ARNOLD
; APPLICANT: BHAVSAR, SUNIL
; TITLE OF INVENTION: USE OF EXTENDINS AND AGONISTS THEREOF FOR
; FILE REFERENCE: 231/181
; CURRENT APPLICATION NUMBER: US 09/003,869A
; EARLIER FILING DATE: 1998-01-07
; EARLIER APPLICATION NUMBER: US 60/034,905
; EARLIER FILING DATE: 1997-01-07
; EARLIER APPLICATION NUMBER: US 60/055,404
; EARLIER FILING DATE: 1997-08-08
; EARLIER APPLICATION NUMBER: US 60/065,442
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: US 60/066,029
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 184
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: artificially synthesized sequence of novel extendin agonist
; OTHER INFORMATION: compound
; FEATURE:
; OTHER INFORMATION: Xaa in positions 31 and 36 stands for homoproline.
; LOCATION: (36)...(36)
; OTHER INFORMATION: amidated hPro (homoprolinamide)
US-09-003-869-184

Query Match          76.7%; Score 92; DB 10; Length 36;
Best Local Similarity 68.8%; Pred. No. 6.4e-10;
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTYXXXXXKQXEEAVRLXXXXXKLNKGXSSGA 35
   || ||| ||||| ||||| |||||
Db 4 GTFTSLSKQMEEEAVRLFTLEWLNKGXSSGA 35
```

```
RESULT 13
US-09-003-869-65
; Sequence 65, Application US/09003869A
; Patent No. US20020137666A1
; GENERAL INFORMATION:
; APPLICANT: BEELEY, NIGEL ROBERT ARNOLD
; APPLICANT: PRICKETT, KATHRYN S.
; APPLICANT: BHAVSAR, SUNIL
; TITLE OF INVENTION: USE OF EXTENDINS AND AGONISTS THEREOF FOR
; FILE REFERENCE: 231/181
; CURRENT APPLICATION NUMBER: US 09/003,869A
; EARLIER FILING DATE: 1998-01-07
; EARLIER APPLICATION NUMBER: US 60/034,905
; EARLIER FILING DATE: 1997-01-07
; EARLIER APPLICATION NUMBER: US 60/055,404
; EARLIER FILING DATE: 1997-08-08
; EARLIER APPLICATION NUMBER: US 60/065,442
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: US 60/066,029
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 65
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: artificially synthesized sequence of novel extendin agonist
; OTHER INFORMATION: compound
; FEATURE:
; NAME/KEY: AMIDATION
; LOCATION: (37)...(37)
; OTHER INFORMATION: amidated Pro (Prolinamide)
US-09-003-869-65

Query Match          76.7%; Score 92; DB 10; Length 37;
Best Local Similarity 65.6%; Pred. No. 6.6e-10;
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTYXXXXXKQXEEAVRLXXXXXKLNKGXSSGA 35
   || ||| ||||| ||||| |||||
Db 4 GTFTSLSKQLEEEAVRLFTLEFLKNGXSSGA 35

RESULT 14
US-09-003-869-83
; Sequence 83, Application US/09003869A
; Patent No. US20020137666A1
; GENERAL INFORMATION:
; APPLICANT: BEELEY, NIGEL ROBERT ARNOLD
; APPLICANT: BHAVSAR, SUNIL
; APPLICANT: PRICKETT, KATHRYN S.
; TITLE OF INVENTION: USE OF EXTENDINS AND AGONISTS THEREOF FOR
; FILE REFERENCE: 231/181
; CURRENT APPLICATION NUMBER: US 09/003,869A
; EARLIER FILING DATE: 1998-01-07
; EARLIER APPLICATION NUMBER: US 60/034,905
; EARLIER FILING DATE: 1997-01-07
; EARLIER APPLICATION NUMBER: US 60/055,404
; EARLIER FILING DATE: 1997-08-08
; EARLIER APPLICATION NUMBER: US 60/065,442
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: US 60/066,029
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 83
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

```

; FEATURE:
; OTHER INFORMATION: artificially synthesized sequence of novel extendin agonist
; OTHER INFORMATION: compound
; FEATURE:
; OTHER INFORMATION: xaa in position 31 stands for n-methylalanine.

```

|                       |              |                    |                |            |
|-----------------------|--------------|--------------------|----------------|------------|
| Query Match           | 76.7%;       | Score 92;          | DB 10;         | Length 37; |
| Best Local Similarity | 68.8%;       | Pred. No. 6.6e-10; |                |            |
| Matches 22:           | Conservative | 0;                 | Mismatches 10; | Indels 0;  |
|                       |              |                    |                | Gaps 0;    |

```

> US-09-003-869-84
> Sequence 84: Application US/09003869A
> Patent No. US20020137666A1
> GENERAL INFORMATION:
> APPLICANT: BEELEY, NIGEL ROBERT ARNOLD
> APPLICANT: BHAVSAR, SUNIL
> APPLICANT: PRICKEIT, KATHRYN S.
> TITLE OF INVENTION: USE OF EXENDINS AND AGONISTS THEREOF FOR
> TITLE OF INVENTION: THE REDUCTION OF FOOD INTAKE
> FILE REFERENCE: 231/181
> CURRENT APPLICATION NUMBER: US/09/003,869A
> CURRENT FILING DATE: 1998-01-07
> EARLIER APPLICATION NUMBER: US 60/034,905
> EARLIER FILING DATE: 1997-01-07
> EARLIER APPLICATION NUMBER: US 60/055,404
> EARLIER FILING DATE: 1997-08-08
> EARLIER APPLICATION NUMBER: US 60/065,442
> EARLIER FILING DATE: 1997-11-14
> EARLIER APPLICATION NUMBER: US 60/066,029
> EARLIER FILING DATE: 1997-11-14
> NUMBER OF SEQ ID NOS: 188
> SOFTWARE: FastSeq for Windows Version 3.0
> SEQ ID NO 84
> LENGTH: 37
> TYPE: PRT
> ORGANISM: Artificial Sequence
> FEATURE:
> OTHER INFORMATION: artificially synthesized sequence of novel exendin agonist
> OTHER INFORMATION: Compound
> FEATURE:
> OTHER INFORMATION: xaa in positions 31, 36 and 37 stands for n-methylalanine.
> FEATURE:
> NAME/KEY: AMIDATION
> LOCATION: (37)...(37)
> OTHER INFORMATION: amidated Nmeala (n-methylalaninamide)
> US-09-003-869-84

```

```
Query Match      76.7%; Score 92; DB 10; Length 37;
Best Local Similarity 68.8%; Pred. No. 6.6e-10;
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

Search completed: January 8, 2003, 11:01:13  
Job time : 8.06034 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2003, 10:58:35 : Search time 216,517 Seconds  
(without alignments)  
116.132 Million cell updates/sec

Title: US-09-003-869-3

Perfect score: 120

Sequence: 1 XXGTXXXXSKXEEAEVRLXXXLXNGXSSGAXXX 39

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

al number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main.\*

- 1: /cgn2\_6/ptodata/1/paa/PCTUS\_COMB.pcp.\*
- 2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pcp.\*
- 3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pcp.\*
- 4: /cgn2\_6/ptodata/1/paa/US08\_COMB.pcp.\*
- 5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pcp.\*
- 6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pcp.\*
- 7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pcp.\*
- 8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pcp.\*
- 9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pcp.\*
- 10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pcp.\*
- 11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pcp.\*
- 12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pcp.\*
- 13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pcp.\*
- 14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pcp.\*
- 15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pcp.\*
- 16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pcp.\*
- 17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pcp.\*
- 18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pcp.\*
- 19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pcp.\*
- 20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pcp.\*
- 21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pcp.\*
- 22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pcp.\*
- 23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pcp.\*
- 24: /cgn2\_6/ptodata/1/paa/US100\_COMB.pcp.\*
- 25: /cgn2\_6/ptodata/1/paa/US101\_COMB.pcp.\*
- 26: /cgn2\_6/ptodata/1/paa/US102\_COMB.pcp.\*
- 27: /cgn2\_6/ptodata/1/paa/US60\_COMB.pcp.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 94    | 78.3        | 39     | 13 | US-08-908-867-35   |
| 2          | 94    | 78.3        | 39     | 13 | US-08-908-867A-35  |
| 3          | 94    | 78.3        | 39     | 13 | US-08-908-867-35   |
| 4          | 93    | 77.5        | 36     | 14 | US-09-003-869-171  |
| 5          | 93    | 77.5        | 36     | 17 | US-09-323-867A-171 |
| 6          | 93    | 77.5        | 36     | 19 | US-09-561-226A-166 |

|    |    |      |    |    |                    |
|----|----|------|----|----|--------------------|
| 7  | 93 | 77.5 | 36 | 19 | US-09-561-226D-166 |
| 8  | 93 | 77.5 | 36 | 21 | US-09-756-690A-171 |
| 9  | 93 | 77.5 | 36 | 22 | US-09-889-331-189  |
| 10 | 93 | 77.5 | 36 | 23 | US-09-554-531A-76  |
| 11 | 93 | 77.5 | 37 | 14 | US-09-003-869-99   |
| 12 | 93 | 77.5 | 37 | 14 | US-09-003-869-183  |
| 13 | 93 | 77.5 | 37 | 17 | US-09-323-867A-99  |
| 14 | 93 | 77.5 | 37 | 17 | US-09-323-867A-183 |
| 15 | 93 | 77.5 | 37 | 19 | US-09-561-226A-86  |
| 16 | 93 | 77.5 | 37 | 19 | US-09-561-226A-178 |
| 17 | 93 | 77.5 | 37 | 19 | US-09-561-226D-86  |
| 18 | 93 | 77.5 | 37 | 19 | US-09-561-226D-178 |
| 19 | 93 | 77.5 | 37 | 20 | US-09-622-105-65   |
| 20 | 93 | 77.5 | 37 | 21 | US-09-756-690A-99  |
| 21 | 93 | 77.5 | 37 | 21 | US-09-756-690A-183 |
| 22 | 93 | 77.5 | 37 | 22 | US-09-889-331-109  |
| 23 | 93 | 77.5 | 37 | 22 | US-09-889-331-201  |
| 24 | 93 | 77.5 | 37 | 23 | US-09-554-531A-88  |
| 25 | 93 | 77.5 | 37 | 23 | US-08-908-867-33   |
| 26 | 93 | 77.5 | 39 | 13 | US-08-908-867A-33  |
| 27 | 93 | 77.5 | 39 | 13 | US-08-908-867-33   |
| 28 | 93 | 77.5 | 39 | 14 | US-09-003-869-35   |
| 29 | 93 | 77.5 | 39 | 14 | US-09-003-869-36   |
| 30 | 93 | 77.5 | 39 | 14 | US-09-003-869-39   |
| 31 | 93 | 77.5 | 39 | 17 | US-09-323-867A-35  |
| 32 | 93 | 77.5 | 39 | 17 | US-09-323-867A-36  |
| 33 | 93 | 77.5 | 39 | 17 | US-09-323-867A-39  |
| 34 | 93 | 77.5 | 39 | 19 | US-09-561-226-36   |
| 35 | 93 | 77.5 | 39 | 19 | US-09-561-226-37   |
| 36 | 93 | 77.5 | 39 | 19 | US-09-561-226-40   |
| 37 | 93 | 77.5 | 39 | 21 | US-09-756-690A-35  |
| 38 | 93 | 77.5 | 39 | 21 | US-09-756-690A-36  |
| 39 | 93 | 77.5 | 39 | 21 | US-09-756-690A-39  |
| 40 | 93 | 77.5 | 39 | 22 | US-09-889-331-36   |
| 41 | 93 | 77.5 | 39 | 22 | US-09-889-331-37   |
| 42 | 93 | 77.5 | 39 | 22 | US-09-889-331-40   |
| 43 | 92 | 76.7 | 35 | 14 | US-09-003-869-69   |
| 44 | 92 | 76.7 | 35 | 14 | US-09-003-869-173  |
| 45 | 92 | 76.7 | 35 | 17 | US-09-323-867A-69  |

ALIGNMENTS

RESULT 1  
US-08-908-867-35  
Sequence 35, Application US/08908867  
GENERAL INFORMATION:  
APPLICANT: YOUNG, Andrew A.  
APPLICANT: GEDULIN, Bronislava  
APPLICANT: BEELEY, Nigel Robert Arnold  
APPLICANT: PRICKETT, Kathryn S.  
TITLE OF INVENTION: METHODS FOR REGULATING  
TITLE OF INVENTION: GASTROINTESTINAL MOTILITY  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LYON & LYON  
STREET: 633 WEST FIFTH STREET  
CITY: LOS ANGELES  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 90017  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/908,867  
FILING DATE: 08-AUGUST-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION NUMBER:  
APPLICATION NUMBER: 08/694,954

SEQUENCE CHARACTERISTICS:  
LENGTH: 39 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
LOCATION: 31, 36, 37, 38  
OTHER INFORMATION: N-methylalanine  
LOCATION: 39  
OTHER INFORMATION: amidated Ser (Serineamide)  
US-08-908-867A-35

Query Match 78.3%; Score 94; DB 13; Length 39;  
Best Local Similarity 65.6%; Pred. No. 3.3e-09;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXKQEEAVRLXXXXLKNKGXSSGA 35  
|| ||| ||||| ||||| |||||  
Db 4 GTFTSLSKQLEEEAVRLFIEFLKNGGASSGA 35

RESULT 3  
US-08-908-867-35  
Sequence 35, Application US/08908867B  
GENERAL INFORMATION:  
APPLICANT: YOUNG, ANDREW A.  
GEDULIN, BRONISLAVA  
BEELEY, NIGEL ROBERT ARNOLD  
PRICKETT, KATHRYN S.  
TITLE OF INVENTION: METHODS FOR REGULATING  
GASTROINTESTINAL MOTILITY  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LYON & LYON  
STREET: 633 WEST FIFTH STREET  
CITY: LOS ANGELES  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 90017  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/908.867B  
FILING DATE: 08-Aug-1997  
CLASSIFICATION: Pending  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/694,954  
FILING DATE: 08-AUGUST-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BERKMAN, CHARLES S.  
REGISTRATION NUMBER: 38,077  
REFERENCE/DOCKET NUMBER: 227/166  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/552-2200  
TELEFAX: 213/955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 35  
SEQUENCE CHARACTERISTICS:  
LENGTH: 39 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
LOCATION: 39  
OTHER INFORMATION: amidated Ser (Serineamide)  
SEQUENCE DESCRIPTION: SEQ ID NO: 35;  
US-08-908-867-35

FILING DATE: 08-AUGUST-1996  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: DUFT, BRADFORD J.  
REGISTRATION NUMBER: 32,219  
REFERENCE/DOCKET NUMBER: 227/166  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/552-2200  
TELEFAX: 213/955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 39 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
LOCATION: 31, 36, 37, 38  
OTHER INFORMATION: N-methylalanine  
LOCATION: 39  
OTHER INFORMATION: amidated Ser (Serineamide)  
US-08-908-867-35

Query Match 78.3%; Score 94; DB 13; Length 39;  
Best Local Similarity 65.6%; Pred. No. 3.3e-09;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXKQEEAVRLXXXXLKNKGXSSGA 35  
|| ||| ||||| ||||| |||||  
Db 4 GTFTSLSKQLEEEAVRLFIEFLKNGGASSGA 35

RESULT 2  
US-08-908-867A-35  
Sequence 35, Application US/08908867A  
GENERAL INFORMATION:  
APPLICANT: YOUNG, ANDREW A.  
GEDULIN, BRONISLAVA  
BEELEY, NIGEL ROBERT ARNOLD  
PRICKETT, KATHRYN S.  
TITLE OF INVENTION: METHODS FOR REGULATING  
GASTROINTESTINAL MOTILITY  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LYON & LYON  
STREET: 633 WEST FIFTH STREET  
CITY: LOS ANGELES  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 90017  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/908.867A  
FILING DATE: 08-AUGUST-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/694,954  
FILING DATE: 08-AUGUST-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: DUFT, BRADFORD J.  
REGISTRATION NUMBER: 32,219  
REFERENCE/DOCKET NUMBER: 227/166  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/552-2200  
TELEFAX: 213/955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 35;

```
Query Match 78.3%; Score 94; DB 13; Length 39;
Best Local Similarity 65.8%; Pred. No. 3.3e-09;
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 4 GTXXXXXSKQEEAEVRLXXXXXKNGXSSGA 35
   || ||| ||||| ||||| ||||| |||||
Db 4 GTFTSDLSKQLEEEAEVRLFTIEFLKNGGSSGA 35

RESULT 4
US-09-003-869-171
; Sequence 171, Application US/09003869A
; GENERAL INFORMATION:
; APPLICANT: BEELEY, NIGEL ROBERT ARNOLD
; APPLICANT: PRICKETT, KATHRYN S.
; APPLICANT: BHAVSAR, SUNIL
; TITLE OF INVENTION: USE OF EXENDINS AND AGONISTS THEREOF FOR
; FILE REFERENCE: 231/181
; CURRENT APPLICATION NUMBER: US/09/003,869A
; PRIOR FILING DATE: 1998-01-07
; EARLIER APPLICATION NUMBER: US 60/034,905
; EARLIER FILING DATE: 1997-01-07
; EARLIER APPLICATION NUMBER: US 60/055,404
; EARLIER FILING DATE: 1997-08-08
; EARLIER APPLICATION NUMBER: US 60/065,442
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: US 60/066,029
; EARLIER FILING DATE: 1997-11-14
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 171
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: artificially synthesized sequence of novel exendin agonist
; OTHER INFORMATION: compound
; NAME/KEY: AMIDATION
; LOCATION: (36)...(36)
; OTHER INFORMATION: amidated Pro (Prolinamide)
US-09-003-869-171

Query Match 77.5%; Score 93; DB 14; Length 36;
Best Local Similarity 65.8%; Pred. No. 4.5e-09;
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 4 GTXXXXXSKQEEAEVRLXXXXXKNGXSSGA 35
   || ||| ||||| ||||| ||||| |||||
Db 4 GTFTSDASKQLEEEAEVRLFTIEFLKNGGPSSGA 35

RESULT 5
US-09-323-867A-171
; Sequence 171, Application US/09323867A
; GENERAL INFORMATION:
; APPLICANT: Amylin Pharmaceuticals, Inc.
; APPLICANT: Young, Andrew et al.
; TITLE OF INVENTION: USE OF EXENDINS AND AGONISTS THEREOF FOR THE TREATMENT
; FILE REFERENCE: 030639.0032.UTL2 (243/131US)
; CURRENT APPLICATION NUMBER: US/09/323,867A
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 189
; SOFTWARE: PatentIn Ver. 2.1 and Microsoft Word
; SEQ ID NO 171
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: artificial sequence with specific variable residues
; NAME/KEY: MOD_RES
```

```
; LOCATION: (36)
; OTHER INFORMATION: AMIDATION, Position 36 is Pro-NH2
US-09-323-867A-171

Query Match 77.5%; Score 93; DB 17; Length 36;
Best Local Similarity 65.6%; Pred. No. 4.5e-09;
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 4 GTXXXXXSKQEEAEVRLXXXXXKNGXSSGA 35
   || ||| ||||| ||||| ||||| |||||
Db 4 GTFTSDASKQLEEEAEVRLFTIEFLKNGGPSSGA 35

RESULT 6
US-09-561-226A-166
; Sequence 166, Application US/09561226A
; GENERAL INFORMATION:
; APPLICANT: Prickett, Kathryn S
; APPLICANT: Young, Andrew A
; TITLE OF INVENTION: MODIFIED EXENDINS AND EXENDIN AGONISTS
; FILE REFERENCE: 030639.0028.UTL(253/204)
; CURRENT APPLICATION NUMBER: US/09/561,226A
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/132,018
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 240
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 166
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Amino Acid Sequence
; NAME/KEY: AMIDATION
; LOCATION: 36
; OTHER INFORMATION: Pro in position 36 is amidated
US-09-561-226A-166

Query Match 77.5%; Score 93; DB 19; Length 36;
Best Local Similarity 65.6%; Pred. No. 4.5e-09;
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 4 GTXXXXXSKQEEAEVRLXXXXXKNGXSSGA 35
   || ||| ||||| ||||| ||||| |||||
Db 4 GTFTSDASKQLEEEAEVRLFTIEFLKNGGPSSGA 35

RESULT 7
US-09-561-226D-166
; Sequence 166, Application US/09561226D
; GENERAL INFORMATION:
; APPLICANT: Prickett, Kathryn S
; APPLICANT: Young, Andrew A
; TITLE OF INVENTION: MODIFIED EXENDINS AND EXENDIN AGONISTS
; FILE REFERENCE: 030639.0028.UTL(253/204)
; CURRENT APPLICATION NUMBER: US/09/561,226D
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/132,018
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 240
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 166
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Amino Acid Sequence
; NAME/KEY: AMIDATION
; LOCATION: 36
; OTHER INFORMATION: Pro in position 36 is amidated
US-09-561-226D-166

Query Match 77.5%; Score 93; DB 19; Length 36;
```



```
; OTHER INFORMATION: compound
; FEATURE:
; OTHER INFORMATION: Xaa in positions 31, 36 and 37 stands for homoproline.
; FEATURE:
; NAME/KEY: AMIDATION
; LOCATION: (37)...(37)
; OTHER INFORMATION: amidated hpro (homoprolinamide)
US-09-003-869-99

Query Match          77.5%; Score 93; DB 14; Length 37;
Best Local Similarity 68.8%; Pred. No. 4.7e-09;
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTXXXXXKQXEEAVRLXXXXXKNGXSSGA 35
   || ||| ||||| ||||| |||||
Db 4 GTFTSDASKQMEEEAVRLFIEWLKNKGXSSGA 35

RESULT 12
US-09-003-869-183
Sequence 183, Application US/09003869A
GENERAL INFORMATION:
; APPLICANT: BEELEY, NIGEL ROBERT ARNOLD
; APPLICANT: PRICKETT, KATHRN S.
; APPLICANT: BHAVSAR, SUNIL
; TITLE OF INVENTION: USE OF EXTENDINS AND AGONISTS THEREOF FOR
; FILE REFERENCE: 231/181
; CURRENT APPLICATION NUMBER: US/09/003,869A
; EARLIER APPLICATION NUMBER: US 60/034,905
; EARLIER FILING DATE: 1998-01-07
; EARLIER APPLICATION NUMBER: US 60/055,404
; EARLIER FILING DATE: 1997-01-07
; EARLIER APPLICATION NUMBER: US 60/065,442
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: US 60/066,029
; EARLIER FILING DATE: 1997-11-14
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 183
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: artificially synthesized sequence of novel extendin agonist
; FEATURE:
; OTHER INFORMATION: Xaa in positions 31, 36 and 37 stands for n-methylalalanine.
; NAME/KEY: AMIDATION
; LOCATION: (37)...(37)
; OTHER INFORMATION: amidated Nmeala (n-methylalalaninamide)
US-09-003-869-183

Query Match          77.5%; Score 93; DB 14; Length 37;
Best Local Similarity 68.8%; Pred. No. 4.7e-09;
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTXXXXXKQXEEAVRLXXXXXKNGXSSGA 35
   || ||| ||||| ||||| |||||
Db 4 GTFTSALSQMEEEAVRLFIEWLKNKGXSSGA 35

RESULT 13
US-09-323-867A-99
Sequence 99, Application US/09323867A
GENERAL INFORMATION:
; APPLICANT: Amylin Pharmaceuticals, Inc.
; APPLICANT: Young, Andrew et al.
; TITLE OF INVENTION: USE OF EXTENDINS AND AGONISTS THEREOF FOR THE TREATMENT
; FILE REFERENCE: 030639.0032.UTL2 (243/131US)
US-09-003-869-99

Query Match          77.5%; Score 93; DB 17; Length 37;
Best Local Similarity 68.8%; Pred. No. 4.7e-09;
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTXXXXXKQXEEAVRLXXXXXKNGXSSGA 35
   || ||| ||||| ||||| |||||
Db 4 GTFTSALSQMEEEAVRLFIEWLKNKGXSSGA 35

RESULT 14
US-09-323-867A-183
Sequence 183, Application US/09323867A
GENERAL INFORMATION:
; APPLICANT: Amylin Pharmaceuticals, Inc.
; APPLICANT: Young, Andrew et al.
; TITLE OF INVENTION: USE OF EXTENDINS AND AGONISTS THEREOF FOR THE TREATMENT
; FILE REFERENCE: 030639.0032.UTL2 (243/131US)
; CURRENT APPLICATION NUMBER: US/09/323,867A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 189
; SOFTWARE: PatentIn Ver. 2.1 and Microsoft Word
; SEQ ID NO 183
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: artificial sequence with specific variable residues
; NAME/KEY: VARIANT
; LOCATION: (31)
; OTHER INFORMATION: Xaa is N-methylalalanine
; NAME/KEY: VARIANT
; LOCATION: (36)...(37)
; OTHER INFORMATION: Xaa is N-methylalalanine
; NAME/KEY: MOD.RES
; LOCATION: (37)
; OTHER INFORMATION: AMIDATION, Position 37 is N-methylalalanine-NH2
US-09-323-867A-183

Query Match          77.5%; Score 93; DB 17; Length 37;
Best Local Similarity 68.8%; Pred. No. 4.7e-09;
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTXXXXXKQXEEAVRLXXXXXKNGXSSGA 35
   || ||| ||||| ||||| |||||
Db 4 GTFTSALSQMEEEAVRLFIEWLKNKGXSSGA 35

RESULT 15
US-09-561-226A-86
Sequence 86, Application US/09561226A
GENERAL INFORMATION:
```

```
; APPLICANT: Prickett, Kathryn S
; APPLICANT: Young, Andrew A
; TITLE OF INVENTION: MODIFIED EXENDINS AND EXENDIN AGONISTS
; FILE REFERENCE: 030639.0028.UTL(253/204)
; CURRENT APPLICATION NUMBER: US/09/561,226A
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/132,018
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 240
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 86
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Amino Acid Sequence
; NAME/KEY: VARIANT
; LOCATION: 31, 36-37
; OTHER INFORMATION: Xaa in positions 31, 36-37 stands for hPro
; NAME/KEY: AMIDATION
; LOCATION: 37
; OTHER INFORMATION: hPro in position 37 is amidated
; US-09-561-226A-86
```

```
Query Match 77.5%; Score 93; DB 19; Length 37;
Best Local Similarity 68.8%; Pred. No. 4.7e-09;
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

```
QY 4 GTXXXXXSKQXEEAVRLXXXXXKNGXSXSGA 35
   |||||
Db 4 GTFTSDASKQMEEEAVRLFIEIENKNGXSXSGA 35
```

```
Search completed: January 8, 2003, 11:12:05
Job time : 217.517 secs
```



```
; Sequence 171, Application US/10157224A
; GENERAL INFORMATION:
; APPLICANT: YOUNG, ANDREW A.
; APPLICANT: KOLTERMAN, ORVILLE G.
; TITLE OF INVENTION: NOVEL EXENDIN AGONIST FORMULATIONS AND METHODS OF
; FILE REFERENCE: 02001-050
; CURRENT APPLICATION NUMBER: US/10/157,224A
; CURRENT FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: 09/889,330
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: PCT/US00/00902
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: 60/116,380
; PRIOR FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: 60/175,365
; PRIOR FILING DATE: 2000-01-10
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn Ver. 2.1
; EQ ID NO 171
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Exendin Agonist
; OTHER INFORMATION: c-term amidation
; US-10-157-224A-171

Query Match 77.5%; Score 93; DB 6; Length 36;
Best Local Similarity 65.8%; Pred. No. 1.8e-10;
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXKQEEAAVRLXXXXLKNKGXSSGA 35
|| ||| ||||| ||||| |||||
Db 4 GTTSDASKQEEAAVRLFIWLNKNGXSSGA 35

RESULT 3
US-10-187-051-99
; Sequence 99, Application US/10187051
; GENERAL INFORMATION:
; APPLICANT: BEELEY, NIGEL ROBERT ARNOLD
; APPLICANT: PRICKETT, KATHRYN S.
; APPLICANT: BHAVSAR, SUNIL
; TITLE OF INVENTION: USE OF EXENDINS AND AGONISTS THEREOF FOR
; FILE REFERENCE: 231/181
; CURRENT APPLICATION NUMBER: US/10/187,051
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US/09/003,869
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: US 60/034,905
; PRIOR FILING DATE: 1997-01-07
; PRIOR APPLICATION NUMBER: US 60/055,404
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: US 60/065,442
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: US 60/066,029
; PRIOR FILING DATE: 1997-11-14
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 99
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: artificially synthesized sequence of novel exendin
; OTHER INFORMATION: agonist
; OTHER INFORMATION: compound
; TITLE OF INVENTION: THE REDUCTION OF FOOD INTAKE
; FILE REFERENCE: 231/181
; CURRENT APPLICATION NUMBER: US/10/187,051
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US/09/003,869
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: US 60/034,905
; PRIOR FILING DATE: 1997-01-07
; PRIOR APPLICATION NUMBER: US 60/055,404
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: US 60/065,442
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: US 60/066,029
; PRIOR FILING DATE: 1997-11-14
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 99
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: artificially synthesized sequence of novel exendin
; OTHER INFORMATION: agonist
; OTHER INFORMATION: compound
; TITLE OF INVENTION: Xaa in positions 31, 36 and 37 stands for homoproline.
; OTHER INFORMATION:
; FILE REFERENCE: 02001-050
```

```
; NAME/KEY: AMIDATION
; LOCATION: (37)...(37)
; OTHER INFORMATION: amidated hpro (homoprolinamide)
; US-10-187-051-99

Query Match 77.5%; Score 93; DB 6; Length 37;
Best Local Similarity 68.8%; Pred. No. 1.9e-10;
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTXXXXXKQEEAAVRLXXXXLKNKGXSSGA 35
|| ||| ||||| ||||| |||||
Db 4 GTTSDASKQEEAAVRLFIWLNKNGXSSGA 35

RESULT 4
US-10-187-051-183
; Sequence 183, Application US/10187051
; GENERAL INFORMATION:
; APPLICANT: BEELEY, NIGEL ROBERT ARNOLD
; APPLICANT: PRICKETT, KATHRYN S.
; APPLICANT: BHAVSAR, SUNIL
; TITLE OF INVENTION: USE OF EXENDINS AND AGONISTS THEREOF FOR
; FILE REFERENCE: 231/181
; CURRENT APPLICATION NUMBER: US/10/187,051
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US/09/003,869
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: US 60/034,905
; PRIOR FILING DATE: 1997-01-07
; PRIOR APPLICATION NUMBER: US 60/055,404
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: US 60/065,442
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: US 60/066,029
; PRIOR FILING DATE: 1997-11-14
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 183
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: artificially synthesized sequence of novel exendin
; OTHER INFORMATION: agonist
; OTHER INFORMATION: compound
; TITLE OF INVENTION: THE REDUCTION OF FOOD INTAKE
; FILE REFERENCE: 231/181
; CURRENT APPLICATION NUMBER: US/10/187,051
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US/09/003,869
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: US 60/034,905
; PRIOR FILING DATE: 1997-01-07
; PRIOR APPLICATION NUMBER: US 60/055,404
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: US 60/065,442
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: US 60/066,029
; PRIOR FILING DATE: 1997-11-14
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 183
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: artificially synthesized sequence of novel exendin
; OTHER INFORMATION: agonist
; OTHER INFORMATION: compound
; TITLE OF INVENTION: Xaa in positions 31, 36 and 37 stands for n-
; OTHER INFORMATION: methylalanine.
; NAME/KEY: AMIDATION
; LOCATION: (37)...(37)
; OTHER INFORMATION: amidated Nmeala (n-methylalaninamide)
; US-10-187-051-183

Query Match 77.5%; Score 93; DB 6; Length 37;
Best Local Similarity 68.8%; Pred. No. 1.9e-10;
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTXXXXXKQEEAAVRLXXXXLKNKGXSSGA 35
|| ||| ||||| ||||| |||||
Db 4 GTTSDASKQEEAAVRLFIWLNKNGXSSGA 35

RESULT 5
US-10-157-224A-99
; Sequence 99, Application US/10157224A
; GENERAL INFORMATION:
; APPLICANT: YOUNG, ANDREW A.
; APPLICANT: KOLTERMAN, ORVILLE G.
; TITLE OF INVENTION: NOVEL EXENDIN AGONIST FORMULATIONS AND METHODS OF
; FILE REFERENCE: 02001-050
; CURRENT APPLICATION NUMBER: US/10/157,224A
; CURRENT FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: 09/889,330
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: PCT/US00/00902
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: 60/116,380
; PRIOR FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: 60/175,365
; PRIOR FILING DATE: 2000-01-10
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn Ver. 2.1
; EQ ID NO 171
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Exendin Agonist
; OTHER INFORMATION: c-term amidation
; US-10-157-224A-171
```



```
; CURRENT APPLICATION NUMBER: US/10/157,224A
; PRIOR FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: 09/889,330
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: PCT/US00/00902
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: 60/116,380
; PRIOR FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: 60/175,365
; PRIOR FILING DATE: 2000-01-10
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 99
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Exendin Agonist
; FEATURE:
; OTHER INFORMATION: c-term amidation
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (31)
; OTHER INFORMATION: Homoproline
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (36)..(37)
; OTHER INFORMATION: Homoproline
US-10-157-224A-99

Query Match          77.5%; Score 93; DB 6; Length 37;
Best Local Similarity 68.8%; Pred. No. 1.9e-10;
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTXXXXXKQEEAEVRLXXXXLKNGXSSGA 35
|| ||| ||||||| |||||||||
Db 4 GTFTSDASKQEEAEVRLFIWLKNGXSSGA 35

RESULT 6
US-10-157-224A-183
; Sequence 183, Application US/10157224A
; GENERAL INFORMATION:
; APPLICANT: YOUNG, ANDREW A.
; TITLE OF INVENTION: NOVEL EXENDIN AGONIST FORMULATIONS AND METHODS OF
; FILE REFERENCE: 02001-050
; CURRENT APPLICATION NUMBER: US/10/157,224A
; PRIOR FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: 09/889,330
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: PCT/US00/00902
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: 60/116,380
; PRIOR FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: 60/175,365
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 183
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Exendin Agonist
; FEATURE:
; OTHER INFORMATION: c-term amidation
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (31)
; OTHER INFORMATION: N-methylalanine
; FEATURE:
```

```
; NAME/KEY: MOD_RES
; LOCATION: (36)..(37)
; OTHER INFORMATION: N-methylalanine
US-10-157-224A-183

Query Match          77.5%; Score 93; DB 6; Length 37;
Best Local Similarity 68.8%; Pred. No. 1.9e-10;
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTXXXXXKQEEAEVRLXXXXLKNGXSSGA 35
|| ||| ||||||| |||||||||
Db 4 GTFTSALSQKEEAEVRLFIWLKNGXSSGA 35

RESULT 7
US-10-187-051-35
; Sequence 35, Application US/10187051
; GENERAL INFORMATION:
; APPLICANT: BEELEY, NIGEL ROBERT ARNOLD
; APPLICANT: PRICKETT, KATHRYN S.
; APPLICANT: BHAVSAR, SUNIL
; TITLE OF INVENTION: USE OF EXENDINS AND AGONISTS THEREOF FOR
; FILE REFERENCE: 231/181
; CURRENT APPLICATION NUMBER: US/10/187,051
; PRIOR FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US/09/003,869
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: US 60/034,905
; PRIOR FILING DATE: 1997-01-07
; PRIOR APPLICATION NUMBER: US 60/055,404
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: US 60/065,442
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: US 60/066,029
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 35
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: artificially synthesized sequence of novel exendin
; OTHER INFORMATION: agonist
; OTHER INFORMATION: compound
; FEATURE:
; OTHER INFORMATION: Xaa in positions 31, 36, 37 and 38 stands for
; OTHER INFORMATION: thiothioline.
; NAME/KEY: AMIDATION
; LOCATION: (39)..(39)
; OTHER INFORMATION: amidated Ser (Serinamide)
US-10-187-051-35

Query Match          77.5%; Score 93; DB 6; Length 39;
Best Local Similarity 68.8%; Pred. No. 2e-10;
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTXXXXXKQEEAEVRLXXXXLKNGXSSGA 35
|| ||| ||||||| |||||||||
Db 4 GTFTSDLSKQEEAEVRLFIWLKNGXSSGA 35

RESULT 8
US-10-187-051-36
; Sequence 36, Application US/10187051
; GENERAL INFORMATION:
; APPLICANT: BEELEY, NIGEL ROBERT ARNOLD
; APPLICANT: PRICKETT, KATHRYN S.
; APPLICANT: BHAVSAR, SUNIL
; TITLE OF INVENTION: USE OF EXENDINS AND AGONISTS THEREOF FOR
; FEATURE:
```



```
; TITLE OF INVENTION: ADMINISTRATION THEREOF
; FILE REFERENCE: 02001-050
; CURRENT APPLICATION NUMBER: US/10/157,224A
; PRIOR FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: 09/889,330
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: PCT/US00/00902
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: 60/116,380
; PRIOR FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: 60/175,365
; PRIOR FILING DATE: 2000-01-10
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Exendin Agonist
; FEATURE:
; OTHER INFORMATION: c-term amidation
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (31)
; OTHER INFORMATION: homoproline
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (36)..(38)
; OTHER INFORMATION: homoproline
; US-10-157-224A-36

Query Match 77.5%; Score 93; DB 6; Length 39;
Best Local Similarity 68.8%; Pred. No. 2e-10;
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTXXXXXKQEEEAVALRLXXXXLKGXSSGA 35
|| ||| ||||| ||||| ||||| |||||
Db 4 GTTDSLKQLEEEAVALRLFIEFLKNGXSSGA 35

RESULT 12
US-10-157-224A-39
; Sequence 39, Application US/10157224A
; GENERAL INFORMATION:
; APPLICANT: KOLTERMAN, ANDREW A.
; TITLE OF INVENTION: NOVEL EXENDIN AGONIST FORMULATIONS AND METHODS OF
; FILE REFERENCE: 02001-050
; CURRENT APPLICATION NUMBER: US/10/157,224A
; PRIOR FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: 09/889,330
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: PCT/US00/00902
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: 60/116,380
; PRIOR FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: 60/175,365
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 39
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Exendin Agonist
; FEATURE:
; OTHER INFORMATION: c-term amidation
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (31)
```

```
; TITLE OF INVENTION: ADMINISTRATION THEREOF
; FILE REFERENCE: 02001-050
; CURRENT APPLICATION NUMBER: US/10/157,224A
; PRIOR FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: 09/889,330
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: PCT/US00/00902
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: 60/116,380
; PRIOR FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: 60/175,365
; PRIOR FILING DATE: 2000-01-10
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Exendin Agonist
; FEATURE:
; OTHER INFORMATION: c-term amidation
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (31)
; OTHER INFORMATION: homoproline
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (36)..(38)
; OTHER INFORMATION: homoproline
; US-10-157-224A-36

Query Match 77.5%; Score 93; DB 6; Length 39;
Best Local Similarity 68.8%; Pred. No. 2e-10;
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 4 GTXXXXXKQEEEAVALRLXXXXLKGXSSGA 35
|| ||| ||||| ||||| ||||| |||||
Db 4 GTTDSLKQLEEEAVALRLFIEFLKNGXSSGA 35

RESULT 13
US-10-187-051-69
; Sequence 69, Application US/10187051
; GENERAL INFORMATION:
; APPLICANT: BEELEY, NIGEL ROBERT ARNOLD
; APPLICANT: PRICKETT, KATHRYN S.
; APPLICANT: BHAVSAR, SUNIL
; TITLE OF INVENTION: USE OF EXENDINS AND AGONISTS THEREOF FOR
; FILE REFERENCE: 231/181
; CURRENT APPLICATION NUMBER: US/10/187,051
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US/09/003,869
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: US 60/034,905
; PRIOR FILING DATE: 1997-01-07
; PRIOR APPLICATION NUMBER: US 60/055,404
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: US 60/065,442
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: US 60/086,029
; PRIOR FILING DATE: 1997-11-14
; NUMBER OF SEQ ID NOS: 188
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 69
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: artificially synthesized sequence of novel exendin
; OTHER INFORMATION: agonist
; OTHER INFORMATION: compound
; FEATURE:
; NAME/KEY: AMIDATION
; LOCATION: (35)...(35)
; OTHER INFORMATION: amidated Ala (Alaninamide)
; US-10-187-051-69

Query Match 76.7%; Score 92; DB 6; Length 35;
Best Local Similarity 65.6%; Pred. No. 2.7e-10;
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXKQEEEAVALRLXXXXLKGXSSGA 35
|| ||| ||||| ||||| ||||| |||||
Db 4 GTTDSLKQLEEEAVALRLFIEFLKNGPSSGA 35

RESULT 14
US-10-187-051-173
; Sequence 173, Application US/10187051
; GENERAL INFORMATION:
; APPLICANT: BEELEY, NIGEL ROBERT ARNOLD
; APPLICANT: PRICKETT, KATHRYN S.
; APPLICANT: BHAVSAR, SUNIL
; TITLE OF INVENTION: USE OF EXENDINS AND AGONISTS THEREOF FOR
; FILE REFERENCE: 231/181
```

Db 4 GTTSDLSKQLEEEAVRLFIEFLKNGPSSGA 35

Search completed: January 8, 2003, 11:12:45  
Job time : 11.0948 secs

; CURRENT APPLICATION NUMBER: US/10/187,051  
; CURRENT FILING DATE: 2002-06-28  
; PRIOR APPLICATION NUMBER: US/09/003,869  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: US 60/034,905  
; PRIOR FILING DATE: 1997-01-07  
; PRIOR APPLICATION NUMBER: US 60/055,404  
; PRIOR FILING DATE: 1997-08-08  
; PRIOR APPLICATION NUMBER: US 60/065,442  
; PRIOR FILING DATE: 1997-11-14  
; PRIOR APPLICATION NUMBER: US 60/066,029  
; PRIOR FILING DATE: 1997-11-14  
; NUMBER OF SEQ ID NOS: 188  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 173  
; LENGTH: 35  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: artificially synthesized sequence of novel extendin  
; OTHER INFORMATION: agonist  
; OTHER INFORMATION: compound  
; FEATURE:  
; NAME/KEY: AMIDATION  
; LOCATION: (35)...(35)  
; OTHER INFORMATION: amidated Ala (Alaninamide)  
US-10-187-051-173

Query Match 76.7%; Score 92; DB 6; Length 35;  
Best Local Similarity 65.6%; Pred. No. 2.7e-10;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXKQLEEEAVRLFIEFLKNGXSSGA 35

Db 4 GTTSDLSKQLEEEAVRLFIEFLKNGPSSGA 35

RESULT 15  
US-10-157-224A-69  
; Sequence 69, Application US/10157224A  
; GENERAL INFORMATION:  
; APPLICANT: YOUNG, ANDREW A.  
; TITLE OF INVENTION: NOVEL EXTENDIN AGONIST FORMULATIONS AND METHODS OF  
; FILE REFERENCE: 02001-050  
; CURRENT APPLICATION NUMBER: US/10/157,224A  
; CURRENT FILING DATE: 2002-05-28  
; PRIOR APPLICATION NUMBER: 09/889,330  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: PCT/US00/00902  
; PRIOR FILING DATE: 2000-01-14  
; PRIOR APPLICATION NUMBER: 60/116,380  
; PRIOR FILING DATE: 1999-01-14  
; PRIOR APPLICATION NUMBER: 60/175,365  
; PRIOR FILING DATE: 2000-01-10  
; NUMBER OF SEQ ID NOS: 188  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 69  
; LENGTH: 35  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Extendin Agonist  
; FEATURE:  
; OTHER INFORMATION: c-term amidation  
US-10-157-224A-69

Query Match 76.7%; Score 92; DB 6; Length 35;  
Best Local Similarity 65.6%; Pred. No. 2.7e-10;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXKQLEEEAVRLFIEFLKNGXSSGA 35

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2003, 10:54:49 ; Search time 12.4397 Seconds  
(without alignments)  
301.394 Million cell updates/sec

Title: US-09-003-869-3

Perfect score: 120

Sequence: 1 XXGTXXXKXQEEAVRLXXXLXNGXSSGAXXXX 39

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

1 number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match % | Length | ID       | Description         |
|------------|-------|---------------|--------|----------|---------------------|
| 1          | 91    | 75.8          | 39     | 1 HWGH32 | extendin-3 - Mexica |
| 2          | 91    | 75.8          | 39     | 1 HWGH4G | extendin-4 - Gila m |
| 3          | 42    | 35.0          | 310    | 2 D86675 | mevalonate kinase   |
| 4          | 42    | 35.0          | 546    | 2 G64803 | phosphoglucosylase  |
| 5          | 42    | 35.0          | 546    | 2 AG0586 | phosphoglucosylase  |
| 6          | 42    | 35.0          | 546    | 2 G85568 | phosphoglucosylase  |
| 7          | 42    | 35.0          | 546    | 2 G90718 | phosphoglucosylase  |
| 8          | 41    | 34.2          | 157    | 2 G75266 | hypothetical prote  |
| 9          | 41    | 34.2          | 357    | 2 T38405 | hypothetical prote  |
| 10         | 41    | 34.2          | 402    | 2 A75054 | molybdenum cofacto  |
| 11         | 41    | 34.2          | 2044   | 2 T13704 | still life protein  |
| 12         | 41    | 34.2          | 2064   | 2 T13707 | still life protein  |
| 13         | 40    | 33.3          | 127    | 2 C69774 | transcription regu  |
| 14         | 40    | 33.3          | 609    | 2 T45637 | beta-D-glucan exoh  |
| 15         | 40    | 33.3          | 772    | 2 T08154 | hypothetical prote  |
| 16         | 39    | 32.5          | 208    | 2 D41076 | probable transcrip  |
| 17         | 39    | 32.5          | 341    | 2 A40706 | extracellular hype  |
| 18         | 39    | 32.5          | 688    | 2 E71845 | polynucleotide      |
| 19         | 39    | 32.5          | 688    | 2 E64671 | polynucleotide pho  |
| 20         | 39    | 32.5          | 1649   | 2 C86822 | hypothetical prote  |
| 21         | 38.5  | 32.1          | 653    | 2 T02080 | probable carbonate  |
| 22         | 38.5  | 32.1          | 1702   | 2 T14050 | protein kinase (PC  |
| 23         | 38    | 31.7          | 272    | 2 A82847 | pyrroline-5-carbox  |
| 24         | 38    | 31.7          | 274    | 2 G97624 | delta 1-pyrrolinase |
| 25         | 38    | 31.7          | 300    | 2 E71023 | probable transcrip  |
| 26         | 38    | 31.7          | 300    | 2 E73110 | transcription init  |
| 27         | 38    | 31.7          | 357    | 2 J4703  | basic helix-loop-h  |
| 28         | 38    | 31.7          | 357    | 2 I49338 | neurogenic differe  |
| 29         | 38    | 31.7          | 381    | 2 A57059 | beta-cell E-box tr  |

30 38 31.7 419 2 S23018 DNA ligase (ATP) (

31 38 31.7 421 2 C85644 hypothetical prote

32 38 31.7 421 2 A90784 hypothetical prote

33 38 31.7 636 2 T45640 beta-D-glucan exoh

34 38 31.7 726 2 T20183 hypothetical prote

35 38 31.7 816 2 D96544 unknown protein [I

36 38 31.7 1464 2 T13716 bazooka gene prote

37 37.5 488 2 C85062 probable thioresox

38 37.5 608 2 D87912 protein B0205.3 [I

39 37 30.8 157 2 B83897 hypothetical prote

40 37 30.8 189 2 G97690 transcribed regu

41 37 30.8 189 2 AD2916 conserved hypothet

42 37 30.8 250 2 AF1095 conserved hypothet

43 37 30.8 250 2 A11458 GTP-binding protei

44 37 30.8 356 2 H90168

45 37 30.8 430 2 S50604 AS22 protein - yea

#### ALIGNMENTS

##### RESULT 1

HWGH32

extendin-3 - Mexican beaded lizard

C:Species: Heloderma horridum (Mexican beaded lizard)

C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 21-Nov-1997

C:Accession: A23674

R:Eng, J.; Andrews, P.C.; Kleinman, W.A.; Singh, L.; Raufman, J.P.

J. Biol. Chem. 265, 20259-20262, 1990

A:Title: Purification and structure of extendin-3, a new pancreatic secretagogue isola

A:Reference number: A23674; MUID:91056067; PMID:1700785

A:Accession: A23674

A:Molecule type: protein

A:Residues: 1-39 <ENG>

C:Comment: Extendins are venom components that are thought to bind to receptors for va

g in secretion of amylase.

C:Superfamily: glucagon

C:Keywords: amidated carboxyl end; duplication; secretagogue; venom

F;39/Modified site: amidated carboxyl end (Ser) #status experimental

Query Match 75.8%; Score 91; DB 1; Length 39;

Best Local Similarity 65.6%; Pred. No. 7.7e-10;

Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXXQXEEAVRLXXXLXNGXSSGA 35

DB 4 GTTSLSKQMEEEAVRLFIEWLXNGGPSSGA 35

##### RESULT 2

HWGH4G

extendin-4 - Gila monster

C:Species: Heloderma suspectum (Gila monster)

C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 21-Nov-1997

C:Accession: A42486

R:Eng, J.; Kleinman, W.A.; Singh, L.; Singh, G.; Raufman, J.P.

J. Biol. Chem. 267, 7402-7405, 1992

A:Title: Isolation and characterization of extendin-4, an extendin-3 analogue, from Hel

A:Reference number: A42486; MUID:92218391; PMID:1313797

A:Accession: A42486

A:Molecule type: protein

A:Residues: 1-39 <ENG>

C:Comment: Extendin-4 does not stimulate amylase secretion by pancreatic acinar cells.

C:Superfamily: glucagon

C:Keywords: amidated carboxyl end; duplication; venom

F;39/Modified site: amidated carboxyl end (Ser) #status experimental

Query Match 75.8%; Score 91; DB 1; Length 39;

Best Local Similarity 65.6%; Pred. No. 7.7e-10;

Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GTXXXXXXQXEEAVRLXXXLXNGXSSGA 35

DB 4 GTTSLSKQMEEEAVRLFIEWLXNGGPSSGA 35











OC Heloderma.  
OX NCBI\_TaxID=8554;  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=97172477; PubMed=9020121;  
RA Chen Y.E., Drucker D.J.;  
RA "Tissue-specific expression of unique mRNAs that encode proglucagon-  
RT derived peptides or exendin 4 in the lizard.";  
RL J. Biol. Chem. 272:4108-4115(1997).  
RN [2]  
RN SEQUENCE OF 48-86.  
RN TISSUE=Venom;  
RC MEDLINE=92218391; PubMed=1313797;  
RX Eng J., Kleinman W.A., Singh L., Singh G., Raufman J.-P.;  
RA "Isolation and characterization of exendin-4, an exendin-3 analogue,  
RT from Heloderma suspectum venom. Further evidence for an exendin  
RT receptor on dispersed acini from guinea pig pancreas.";  
RL J. Biol. Chem. 267:7402-7403(1992).  
RN [3]  
RN FUNCTION: HAS A VIP/SECRETIN-LIKE BIOLOGICAL ACTIVITY. INTERACTS  
CC WITH THE EXENDIN RECEPTOR.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- TISSUE SPECIFICITY: Produced by the venomous gland.  
CC -!- SIMILARITY: BELONGS TO THE GLUCAGON FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U77613; AAB51130.1; -;  
DR PIR; A42486; HMGH4G.  
DR InterPro; IPR000532; Glucagon.  
DR Pfam; PF00123; hormone2; 1.  
DR SMART; SM00070; GLUCA; 1.  
DR PROSITE; PS00260; GLUCAGON; 1.  
KW Glucagon family; Toxin; Amidation; Signal.  
FT SIGNAL 1 23  
FT PEPTIDE 48 86  
FT MOD\_RES 86 86  
FT SEQUENCE 87 AA; 9479 MW; 656BA6E3D874542 CRC64;  
Query Match 75.8%; Score 91; DB 1; Length 87;  
Best Local Similarity 65.6%; Pred. No. 1.8e-09;  
Matches 21; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
4 GTXXXXXSKXEEAVRLXXXXLKNKGXSSGA 35  
51 GTFTSDLSKQMEAEAVRLFTEWLKNGPSSGA 82  
Db  
RESULT 3  
PGMU\_ECOLI  
ID PGMU\_ECOLI STANDARD; PRT; 546 AA.  
AC P36938;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Phosphoglucosyltransferase (EC 5.4.2.2) (Glucose phosphomutase) (PGM).  
GN PGM OR B0688.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia  
OX NCBI\_TaxID=562;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=94364967; PubMed=8083177;  
RA Lu M., Kleckner N.;  
RT "Molecular cloning and characterization of the pgm gene encoding  
RT phosphoglucosyltransferase of Escherichia coli.";

RL J. Bacteriol. 176:5847-5851(1994).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [3]  
RN SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horiuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 12.7-28.0 min region on the linkage map.";  
RL DNA Res. 3:137-155(1996).  
RN [4]  
RN SEQUENCE OF 1-20 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=94236686; PubMed=8011018;  
RA Lu M., Campbell J.L., Boye E., Kleckner N.;  
RT "SeqA: a negative modulator of replication initiation in E. coli.";  
RL Cell 77:413-426(1994).  
RN [5]  
RN CHARACTERIZATION.  
RP Joshi J.G., Handler P.;  
RA "Phosphoglucosyltransferase. II. Purification and properties of  
RT phosphoglucosyltransferase from Escherichia coli.";  
RL J. Biol. Chem. 239:2741-2751(1964).  
CC -!- FUNCTION: THIS ENZYME PARTICIPATES IN BOTH THE BREAKDOWN AND  
CC SYNTHESIS OF GLUCOSE.  
CC -!- CATALYTIC ACTIVITY: Alpha-D-glucose 1-phosphate = alpha-D-glucose  
CC 6-phosphate.  
CC -!- SIMILARITY: BELONGS TO THE PHOSPHOHEXOSE MUTASES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U08369; AAA57067.1; -;  
DR EMBL; AE000172; AAC73782.1; -;  
DR EMBL; D90707; BAA35337.1; -;  
DR EMBL; D90708; BAA35345.1; -;  
DR EMBL; U07651; -; NOT\_ANNOTATED\_CDS.  
DR EcoGene; EG12144; pgm.  
DR InterPro; IPR001485; PG/PMM\_mutase.  
DR Pfam; PF00408; PGM\_PMM; 1.  
DR Pfam; PF02878; PGM\_PMM\_I; 1.  
DR Pfam; PF02879; PGM\_PMM\_II; 1.  
DR Pfam; PF02880; PGM\_PMM\_III; 1.  
DR TIGRFAMs; TIGR01132; pgm; 1.  
DR PROSITE; PS00710; PGM\_PMM; 1.  
KW Isomerase; Phosphorylation; Complete proteome.  
FT ACT\_SITE 145 146  
FT FORMS THE PHOSPHOSERINE INTERMEDIATE  
FT (BY SIMILARITY).  
SQ SEQUENCE 546 AA; 58361 MW; 666B6B9C2F2ECD59 CRC64;  
Query Match 35.0%; Score 42; DB 1; Length 546;  
Best Local Similarity 52.9%; Pred. No. 6.5;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 12 KOXEEAVRLXXXXLKN 28  
 II : III : III  
 Db 529 KOIEKEAVEIVSEVLKN 545

## RESULT 4

YFQB\_SCHPO STANDARD; PRT; 357 AA.  
 AC YFQB\_SCHPO STANDARD; PRT; 357 AA.

DT 01-OCT-1996 (Rel. 34, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypothetical protein c8E11.11 in chromosome I.

GN SPAC8E11.11 OR SPAC26A3.17C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI\_TaxID=4896;

BN [1]

SEQUENCE FROM N.A.

STRAIN=972;

MDLINE=21848401; PubMed=11859360;

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,

RA Sproules J., Feat N., Hayes J., Baker S., Basham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,

RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neill S., Pearson D., Quay M.A., Rabinowitsch E.,

RA Rutherford K., Butter S., Saunders D., Seeger K., Sharp S.,

RA Skelton J., Simmonds M., Squares R., Stevens K.,

RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Volkert G., Aert R., Robben J., Grymonprez B.,

RA Welljens I., Vantreest E., Rieger M., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,

RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,

RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,

RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

RA Lucas M., Rochet M., Gaillard C., Tallada V.A., Garzon A., Thode G.,

RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,

RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,

RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,

RT "The genome sequence of Schizosaccharomyces pombe."

Nature 415:871-880(2002).

-!- SIMILARITY: SOME, TO RAT GUANIDINOACETATE N-METHYLTRANSFERASE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AL021817; CAB40198.1; -

DR EMBL; 269240; CAA93240.1; -

DR InterPro; IPR002110; ANK.

KW Hypothetical protein.

SQ SEQUENCE 357 AA; 40709 MW; 5529B8D3B88BD91A9 CRC64;

Query Match 34.2%; Score 41; DB 1; Length 357;  
 Best Local Similarity 34.8%; Pred. No. 6.3;  
 Matches 8; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

OY 12 KOXEEAVRLXXXXLKNXSG 34

II : III : III

Db 64 KETEVAIEVTKWILNSGVVNG 86

RESULT 5  
 ENCL\_HUMAN STANDARD; PRT; 589 AA.  
 ID ENCL\_HUMAN STANDARD; PRT; 589 AA.  
 AC 014662; Q9UPG9; O75464;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ectoderm-neural cortex-1 protein (ENC-1) (p53-induced protein 10)  
 DE (Nuclear matrix protein NRP/B).  
 GN ENCL OR PIG10 OR NRPB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Colon cancer;  
 RX MEDLINE=97449378; PubMed=9305847;  
 RA Polyak K., Xia Y., Zweier J.L., Kinzler K.W., Vogelstein B.;  
 RT "A model for p53-induced apoptosis.";  
 RL Nature 389:300-305(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=98350113; PubMed=9883534;  
 RA Hernandez M.-C., Andres-Barquin P.J., Holt I., Israel M.A.;  
 RT "Cloning of human ENC-1 and evaluation of its expression and  
 RT regulation in nervous system tumors.";  
 RL Exp. Cell Res. 242:470-477(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Hippocampus, and Fetal brain;  
 RX MEDLINE=98234394; PubMed=9566959;  
 RA Kim T.-A., Lim J., Ota S., Raja S., Rogers R., Rivnay B., Avraham H.,  
 RA Avraham S.;  
 RT "NRP/B, a novel nuclear matrix protein, associates with p110(RB) and  
 RT is involved in neuronal differentiation.";  
 RL J. Cell Biol. 141:553-566(1998).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Muscle;  
 RA Strausberg R.;  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: ACTIN-BINDING PROTEIN INVOLVED IN THE REGULATION OF  
 CC NEURONAL PROCESS FORMATION AND IN DIFFERENTIATION OF NEURAL CREST  
 CC CELLS. MAY BE DOWN-REGULATED IN NEUROBLASTOMA TUMORS.  
 CC -!- SUBUNIT: BINDS TO RBI. HYPOPHOSPHORYLATED RBI ASSOCIATES WITH ENCL  
 CC DURING NEURONAL DIFFERENTIATION, WHILE HYPERPHOSPHORYLATED RBI  
 CC ASSOCIATES WITH ENCL IN NONDIFFERENTIATING CELLS.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR. NUCLEAR MATRIX-ASSOCIATED.  
 CC -!- TISSUE SPECIFICITY: DETECTED IN FETAL BRAIN TISSUE, MODERATE IN  
 CC ADULT BRAIN, PARTICULARLY HIGH IN THE HIPPOCAMPUS AND  
 CC AMYGDALA, AND SPINAL CHORD. DETECTABLE IN ADULT PANCREAS.  
 CC -!- DEVELOPMENTAL STAGE: DRAMATICALLY UPREGULATED UPON NEURONAL  
 CC DIFFERENTIATION.  
 CC -!- PTM: PHOSPHORYLATED.  
 CC -!- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 6 KELCH REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).



```

CC      SEGMENT OF THE CENTRAL NERVOUS SYSTEM. AT STAGE 17, EXPRESSION
CC      BECOMES RESTRICTED TO THE SYNAPTIC REGIONS OF THE BRAIN AND
CC      VENTRAL NERVE CORD, WHERE SYNAPSES UNDERGO MATURATION.
CC      -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.
CC      -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC      -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: D66546; BAA13108.1; -.
CC      HSSP: P08567; LPLS.
CC      FlyBase: FBgn0019652; sif.
CC      InterPro: IPR001331; GDS_CDC24.
CC      InterPro: IPR001478; PDZ.
CC      InterPro: IPR001849; PH.
CC      InterPro: IPR003116; RBD.
CC      InterPro: IPR000219; RhoGEF.
CC      Pfam: PF00169; PH; 2.
CC      Pfam: PF00621; RhoGEF; 1.
CC      Pfam: PF02196; RBD; 1.
CC      SMART: SM00228; PDZ; 1.
CC      SMART: SM00233; PH; 2.
CC      SMART: SM00455; RBD; 1.
CC      SMART: SM00325; RhoGEF; 1.
CC      PROSITE: PS0010; DH_2; 1.
CC      PROSITE: PS00741; DH_1; 1.
CC      PROSITE: PS50106; PDZ; 1.
CC      PROSITE: PS50003; PH_DOMAIN; 1.
CC      PROSITE: PS50003; PH_DOMAIN; 1.
CC      Guanine-nucleotide releasing factor; Developmental protein; Synapse;
CC      Repeat; Alternative splicing.
KW      Repeat: 62 249 4 X 25 AA APPROXIMATE REPEAT.
FT      DOMAIN 62 249
FT      REPEAT 62 86 1.
FT      REPEAT 94 118 2.
FT      REPEAT 154 178 3.
FT      REPEAT 225 249 4.
FT      DOMAIN 819 937 PH 1.
FT      DOMAIN 1184 1273 PDZ.
FT      DOMAIN 1408 1602 DH.
FT      DOMAIN 1674 1767 PH 2.
FT      DOMAIN 467 470 POLY-PRO.
FT      DOMAIN 646 649 POLY-ARG.
FT      DOMAIN 1295 1298 POLY-PRO.
FT      DOMAIN 1898 1909 POLY-GLN.
FT      DOMAIN 1929 1933 POLY-PRO.
SU      SEQUENCE 2044 AA; 228324 MW; 75D7CF21F49654B6 CRC64;

Query Match          34.2%; Score 41; DB 1; Length 2044;
Best Local Similarity 41.7%; Pred. No. 38;
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 12 KQEEAEVRLXXXXXKNGXSSGA 35
DB 1761 RQIRESVRNMSIPMKNFGGSSGS 1784

RESULT 8
SIFL_DROME
ID SIFL_DROME STANDARD; PRT: 2064 AA.
AC P91621;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Still life protein type 1 (SIF type 1).
GN SIF.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

```

```

OC      Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OC      NCBI_TaxID=7227;
OX [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Head;
RC MEDLINE=97153054; PubMed=8999801;
RA Sone M., Hoshino M., Suzuki E., Kuroda S., Kaibuchi K., Nakagoshi H.,
RA Saigo K., Nabeshima Y.-I., Hama C.;
RT "Still life, a protein in synaptic terminals of Drosophila homologous
RT to GDP-GTP exchangers";
RL Science 275:543-547(1997).
RN ERRATUM.
RP Sone M., Hoshino M., Suzuki E., Kuroda S., Kaibuchi K., Nakagoshi H.,
RA Sone M., Nabeshima Y.-I., Hama C.;
RL Science 275:1405-1405(1997).
CC      -!- FUNCTION: REGULATES SYNAPTIC DIFFERENTIATION THROUGH THE
CC      ORGANIZATION OF ACTIN CYTOSKELETON POSSIBLY BY ACTIVATING RHO-LIKE
CC      GTPASES. IS LIKELY A FACTOR IN THE CASCADE OF RAC1 OR CDC42 IN THE
CC      NEURONS.
CC      -!- SUBCELLULAR LOCATION: LOCALIZES TO THE SUBMEMBRANOUS REGION OF
CC      SYNAPTIC TERMINALS.
CC      -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SIF TYPE 1 (SHOWN HERE) AND SIF
CC      TYPE 2 (AC P91620); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC      -!- DEVELOPMENTAL STAGE: AT STAGE 14, EXPRESSION OCCURS IN EACH
CC      SEGMENT OF THE CENTRAL NERVOUS SYSTEM. AT STAGE 17, EXPRESSION
CC      BECOMES RESTRICTED TO THE SYNAPTIC REGIONS OF THE BRAIN AND
CC      VENTRAL NERVE CORD, WHERE SYNAPSES UNDERGO MATURATION.
CC      -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.
CC      -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC      -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: D66547; BAA13109.1; -.
CC      HSSP: P08567; LPLS.
CC      FlyBase: FBgn0019652; sif.
CC      InterPro: IPR001331; GDS_CDC24.
CC      InterPro: IPR001478; PDZ.
CC      InterPro: IPR001849; PH.
CC      InterPro: IPR003116; RBD.
CC      InterPro: IPR000219; RhoGEF.
CC      InterPro: IPR001960; WHI.
CC      Pfam: PF00169; PH; 2.
CC      Pfam: PF00621; RhoGEF; 1.
CC      Pfam: PF02196; RBD; 1.
CC      SMART: SM00228; PDZ; 1.
CC      SMART: SM00233; PH; 2.
CC      SMART: SM00455; RBD; 1.
CC      SMART: SM00325; RhoGEF; 1.
CC      SMART: SM00461; WHI; 1.
CC      PROSITE: PS50010; DH_2; 1.
CC      PROSITE: PS00741; DH_1; 1.
CC      PROSITE: PS50106; PDZ; 1.
CC      PROSITE: PS50003; PH_DOMAIN; 1.
CC      Guanine-nucleotide releasing factor; Developmental protein; Repeat;
KW      Myristate; Synapse; Alternative splicing.
FT      LIPID 2 2 MYRISTATE (POTENTIAL).
FT      DOMAIN 839 957 PH 1.
FT      DOMAIN 1204 1293 PDZ.
FT      DOMAIN 1428 1622 DH.
FT      DOMAIN 1694 1787 PH 2.
FT      DOMAIN 445 453 POLY-GLY.
FT      DOMAIN 545 548 POLY-GLN.
FT      DOMAIN 1315 1318 POLY-PRO.
FT      DOMAIN 1918 1929 POLY-GLN.
FT      DOMAIN 1949 1953 POLY-PRO.

```

```

AC P35674;
AC 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Harpin-PSS.
GN HRP2.
OS Pseudomonas syringae (pv. syringae).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OC NCBI_TaxID=321;
RN [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 141-162.
RC STRAIN=61;
RC MEDLINE=93313957; Pubmed=8324821;
RA He S.Y., Huang H.-C., Collmer A.;
RT "Pseudomonas syringae pv. syringae harpinss: a protein that is
RT secreted via the hrp pathway and elicits the hypersensitive response
RT in plants.";
RL Cell 73:1255-1266(1993).
CC -!- FUNCTION: ELICITS THE HYPERSENSITIVE RESPONSE (HR) IN THE PLANT
CC UPON INFECTION. HARPIN ELICITS HR IN NON-HOSTS AND IS ALSO
CC REQUIRED FOR PATHOGENICITY IN HOST PLANTS.
CC -!- SUBCELLULAR LOCATION: SECRETED; VIA THE HRP SECRETION PATHWAY.
CC -!- MISCELLANEOUS: DIFFERENT PLANTS EXHIBIT DIFFERENT LEVELS OF
CC SENSITIVITY TO HARPIN-PSS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L14775; AAA25839.1;
CC FIR; A40706; A40706.
CC Hypersensitive response; Repeat. 2 X 7 AA REPEATS OF G-G-G-L-G-T-P.
KW DOMAIN 210 271 1-1.
FT REPEAT 210 216 1-1.
FT REPEAT 265 271 1-2.
FT DOMAIN 276 314 2 X 4 AA REPEATS OF Q-T-G-T.
FT REPEAT 276 279 2-1.
FT REPEAT 311 314 2-2.
FT SEQUENCE 341 AA; 34721 MW; 75FB7329B5380179 CRC64;
SQ
Query Match 32.5%; Score 39; DB 1; Length 341;
Best Local Similarity 32.0%; Pred. No. 14;
Matches 8; Conservative 4; Mismatches 13; Indels 0; Gaps 0;
QY 5 TXXXXSKQXEAEAVRLXXXLKGNG 29
DB 27 TTGSTSSKALQEVVVKLBELMRNG 51
RESULT 11
TF2B_PYRAB STANDARD; PRT; 300 AA.
AC Q9V0V5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription initiation factor IIB (TFIIB).
GN TFB OR PAB1912.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OC NCBI_TaxID=29292;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=GE5 / Orsay;
RA Helling R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution.";

```

```

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC
CC -!- FUNCTION: STABILIZES TBP BINDING TO AN ARCHAEAL BOX-A PROMOTER.
CC ALSO RESPONSIBLE FOR RECRUITING RNA POLYMERASE II TO THE PRE-
CC INITIATION COMPLEX (DNA-TBP-TFIIB) (BY SIMILARITY).
CC
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC
CC -!- SIMILARITY: BELONGS TO THE TFIIB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ248285; CAB49598.1; -.
CC HSSP: P29095; LAIS.
CC InterPro: IPR004366; Cyclin.
CC InterPro: IPR000812; TFIIB_euk.
CC Pfam: PF00382; transcript_fac2; 2.
CC PRINTS: PR00685; TIFACTORIIB.
CC SMART: SM00385; CYCLIN; 2.
CC PROSITE: PS00782; TFIIB; 2.
CC Transcription regulation; Repeat; Zinc-finger; Metal-binding; Zinc;
CC Complete proteome.
CC ZN_FING 7 29 ZN-RIBBON TFIIB-TYPE.
CC REPEAT 114 197 1.
CC REPEAT 210 291 2.
CC METAL 7 7 ZINC (BY SIMILARITY).
CC METAL 10 10 ZINC (BY SIMILARITY).
CC METAL 26 26 ZINC (BY SIMILARITY).
CC METAL 29 29 ZINC (BY SIMILARITY).
CC SEQUENCE 300 AA; 34069 MW; D7AE15181A36BDAF CRC64;
CC
CC Query Match 31.7%; Score 38; DB 1; Length 300;
CC Best Local Similarity 44.4%; Pred. No. 18;
CC Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;
CC
CC QY 12 KXEEEAARLXXXXLKG 29
CC | |||| || : : |
CC Db 127 KHVEEAARLYREAVRKG 144
CC
CC RESULT 12
CC ID TF2B_PYRHO STANDARD; PRT; 300 AA.
CC AC O59151;
CC DT 30-MAY-2000 (Rel. 39, Created)
CC 30-MAY-2000 (Rel. 39, Last sequence update)
CC 16-OCT-2001 (Rel. 40, Last annotation update)
CC Transcription initiation factor IIB (TFIIB).
CC TFB OR PH1482.
CC GN Pyrococcus horikoshii.
CC OS Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
CC OC Pyrococcus.
CC OX NCBI_TaxID=53953;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=OT3;
CC RX MEDLINE=98344137; PubMed=9679194;
CC RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
CC Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
CC Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
CC Funahashi T., Tanaka T., Kudo H., Yamazaki J., Kishida N., Oguchi A.,
CC Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
CC Masuchi Y., Shizuya H., Kikuchi H.;
CC "Complete sequence and gene organization of the genome of a hyper-
CC thermophilic archaeobacterium, Pyrococcus horikoshii OT3";
CC DNA Res. 5:55-76(1998).
CC
CC -!- FUNCTION: STABILIZES TBP BINDING TO AN ARCHAEAL BOX-A PROMOTER.
CC ALSO RESPONSIBLE FOR RECRUITING RNA POLYMERASE II TO THE PRE-
CC INITIATION COMPLEX (DNA-TBP-TFIIB) (BY SIMILARITY).
CC
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).

```

```

CC
CC -!- SIMILARITY: BELONGS TO THE TFIIB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AP000006; BAA30589.1; -.
CC HSSP: P29095; LAIS.
CC InterPro: IPR004366; Cyclin.
CC InterPro: IPR000812; TFIIB_euk.
CC Pfam: PF00382; transcript_fac2; 2.
CC PRINTS: PR00685; TIFACTORIIB.
CC SMART: SM00385; CYCLIN; 2.
CC PROSITE: PS00782; TFIIB; 2.
CC Transcription regulation; Repeat; Zinc-finger; Metal-binding; Zinc;
CC Complete proteome.
CC ZN_FING 7 29 ZN-RIBBON TFIIB-TYPE.
CC REPEAT 114 197 1.
CC REPEAT 210 291 2.
CC METAL 7 7 ZINC (BY SIMILARITY).
CC METAL 10 10 ZINC (BY SIMILARITY).
CC METAL 26 26 ZINC (BY SIMILARITY).
CC METAL 29 29 ZINC (BY SIMILARITY).
CC SEQUENCE 300 AA; 34097 MW; DE9758F398BC855F CRC64;
CC
CC Query Match 31.7%; Score 38; DB 1; Length 300;
CC Best Local Similarity 44.4%; Pred. No. 18;
CC Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;
CC
CC QY 12 KXEEEAARLXXXXLKG 29
CC | |||| || : : |
CC Db 127 KHVEEAARLYREAVRKG 144
CC
CC RESULT 13
CC ID NDF1_MESAU STANDARD; PRT; 355 AA.
CC AC Q60430;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC 01-NOV-1997 (Rel. 35, Last sequence update)
CC 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE Neurogenic differentiation factor 1 (NeuroD1) (Beta-cell E-box trans-
CC activator 2) (BETA2).
CC GN NEUROD1 OR NEUROD.
CC OS Mesocricetus auratus (Golden hamster).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
CC Mesocricetus.
CC OX NCBI_TaxID=10036;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=95293222; PubMed=7774807;
CC Naya F.J., Steilrecht C.M.M., Tsai M.-J.;
CC "Tissue-specific regulation of the insulin gene by a novel basic
CC helix-loop-helix transcription factor.";
CC Genes Dev. 9:1009-1019(1995).
CC
CC -!- FUNCTION: ACTS AS A DIFFERENTIATION FACTOR DURING NEUROGENESIS.
CC TRANSCRIPTIONAL ACTIVATOR. BINDS TO THE INSULIN GENE E-BOX.
CC
CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. HETERODIMER WITH E47.
CC
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC
CC -!- TISSUE SPECIFICITY: MOST ABUNDANT IN PANCREATIC ALPHA- AND BETA-
CC CELLS, LESS IN BRAIN AND INTESTINE.
CC
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U24679; AAA86518.1; ALT_INIT.
DR InterPro: IPR001092; HLH_basic.
DR Pfam: PF00010; HLH; 1.
DR SMART: SM00353; HLH; 1.
DR PROSITE: PS00038; HLH_1; 1.
DR PROSITE: PS00888; HLH_2; 1.
DR DNA-binding; Nuclear protein; Transcription regulation; Activator;
KW Neurogenesis; Developmental protein; Differentiation.
FT DOMAIN 58 77
FT DOMAIN 86 92
FT DNA_BIND 101 112
FT DOMAIN 113 153
FT DOMAIN 167 175
FT DOMAIN 186 194
FT POLY-LYS.
SQ SEQUENCE 355 AA; 39763 MW; F4344DFD360226B2 CRC64;
Query Match 31.7%; Score 38; DB 1; Length 355;
Best Local Similarity 42.1%; Pred. No. 21;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
QY 12 KQEEEAARVRLXXXXLKNKG 30
DB 39 KEDELEAMNAEDSLRNGG 57
-----
RESULT 14
NDFL_MOUSE
ID NDFL_MOUSE STANDARD; PRT; 357 AA.
AC Q60867; Q60897;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic differentiation factor 1 (NeuroD1).
DE NEUROD1 OR NEUROD.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MF1, and 129/SV;
RX MEDLINE=95273957; PubMed=7754368;
Lee J.E., Hollenberg S.M., Snider L., Turner D.L., Lipnick N.,
Weintraub H.;
"Conversion of Xenopus ectoderm into neurons by NeuroD, a basic
helix-loop-helix protein";
Science 268:836-844(1995).
CC -!- FUNCTION: ACTS AS A DIFFERENTIATION FACTOR DURING NEUROGENESIS.
CC TRANSCRIPTIONAL ACTIVATOR. BINDS TO THE INSULIN GENE E-BOX.
CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. HETERODIMER WITH E47.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN DIFFERENTIATING NEURONS OF
CC BOTH THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYONIC DEVELOPMENT.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: U28068; AAC52203.1; -.
DR EMBL: U28888; AAC52204.1; -.
-----
DR MGI:1339708; Neurod1.
DR InterPro: IPR001092; HLH_basic.
DR Pfam: PF00010; HLH; 1.
DR SMART: SM00353; HLH; 1.
DR PROSITE: PS00038; HLH_1; 1.
DR PROSITE: PS00888; HLH_2; 1.
DR DNA-binding; Nuclear protein; Transcription regulation; Activator;
KW Neurogenesis; Developmental protein; Differentiation.
FT DOMAIN 58 77
FT DOMAIN 86 92
FT DNA_BIND 102 113
FT DOMAIN 114 134
FT DOMAIN 158 164
FT POLY-GLU.
FT POLY-GLU.
FT POLY-LYS.
SQ SEQUENCE 357 AA; 39998 MW; B6626E1315E31027 CRC64;
Query Match 31.7%; Score 38; DB 1; Length 357;
Best Local Similarity 42.1%; Pred. No. 22;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
QY 12 KQEEEAARVRLXXXXLKNKG 30
DB 39 KEDELEAMNAEDSLRNGG 57
-----
RESULT 15
NDFL_RAT
ID NDFL_RAT STANDARD; PRT; 357 AA.
AC Q64289;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic differentiation factor 1 (NeuroD1).
DE factor 1) (BHF-1).
GN NEUROD1 OR NEUROD.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Corebellum;
RX MEDLINE=96220182; PubMed=8660336;
Kawakami H., Maruyama H., Yasunami M., Ohkubo H., Hara H., Saïda T.,
Nakanishi S., Nakamura S.;
"Cloning and expression of a rat brain basic helix-loop-helix
factor.";
Biochem. Biophys. Res. Commun. 221:199-204(1996).
RN [2]
RP SEQUENCE OF 88-200 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Retina;
Ahmad I., Acharay H.R.;
Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACTS AS A DIFFERENTIATION FACTOR DURING NEUROGENESIS.
CC TRANSCRIPTIONAL ACTIVATOR. BINDS TO THE INSULIN GENE E-BOX.
CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. HETERODIMER WITH E47.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: D82075; BAA11536.1; -.
DR EMBL: D82074; BAA11535.1; -.
DR EMBL: U80603; AAB38744.1; -.
-----
```



DR InterPro: IPR001092; HLH\_basic.  
DR Pfam: PF00010; HLH; 1.  
DR SMART: SM00353; HLH; 1.  
DR PROSITE: PS00038; HLH\_1; 1.  
DR PROSITE: PS50888; HLH\_2; 1.  
KW DNA-binding; Nuclear protein; Transcription regulation; Activator;  
KW Neurogenesis; Developmental protein; Differentiation.  
FT DOMAIN 58 77  
FT DOMAIN 87 93 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DNA\_BIND 102 113 BASIC DOMAIN.  
FT DOMAIN 114 154 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
FT DOMAIN 67 76 POLY-GLU.  
FT DOMAIN 87 90 POLY-LYS.  
SQ SEQUENCE 357 AA; 40000 MW; F773637E64D3E99E CRC64;  
  
Query Match 31.7%; Score 38; DB 1; Length 357;  
Best Local Similarity 42.1%; Pred. No. 22;  
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;  
  
12 KQEEEEAVRLXXXXLKNGG 30  
1:111:111  
39 KEDELEAMNAEEDSLRNGG 57

Search completed: January 8, 2003, 10:58:50  
Job time : 9.72414 secs





OS Herpes simplex virus (type 2).  
OC Viruses; dsDNA viruses, no RNA stage: Herpesviridae;  
OC Alphaherpesvirinae: Simplexvirus.  
OX NCBI\_TaxID=10310;

RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=HG52;

RX MEDLINE=87111457; PubMed=3027242;

RA McGeoch D.J., Moss H.W., McNab D., Frame M.C.;

RT "DNA sequence and genetic content of the HindIII 1 region in the short  
unique component of the herpes simplex virus type 2 genome;

RT identification of the gene encoding glycoprotein G, and evolutionary  
comparisons.";

RL J. Gen. Virol. 68:19-38(1987).

RN [2]  
RP SEQUENCE FROM N.A.

RC STRAIN=HG52;

RX MEDLINE=90278430; PubMed=2161906;

RA Everett R., Fenwick M.;

RT "Comparative DNA sequence analysis of the host shutoff genes of  
different strains of herpes simplex virus: type 2 strain HG52 encodes

a truncated UL41 product.";

RL J. Gen. Virol. 71:1387-1390(1990).

RN [3]  
RP SEQUENCE FROM N.A.

RC STRAIN=HG52;

RX MEDLINE=92113549; PubMed=1662697;

RA McGeoch D.J., Cunningham C., McIntyre G., Dolan A.;

RT "Comparative sequence analysis of the long repeat regions and  
adjoining parts of the long unique regions in the genomes of herpes

simplex viruses types 1 and 2.";

RL J. Gen. Virol. 72:3057-3075(1991).

RN [4]  
RP SEQUENCE FROM N.A.

RC STRAIN=HG52;

RX MEDLINE=92356101; PubMed=1322965;

RA Barnett B.C., Dolan A., Telford E.A.R., Davison A.J., McGeoch D.J.;

RT "A novel herpes simplex virus gene (UL49A) encodes a putative membrane  
protein with counterparts in other herpesviruses.";

RL J. Gen. Virol. 73:2167-2171(1992).

RN [5]  
RP SEQUENCE FROM N.A.

RC STRAIN=HG52;

RA Dolan A.;

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: Z86099; CAB06753.1; -.

DR InterPro: IPR000501; Proc\_transprot.

DR Pfam: PF01366; PRTF; 1.

SEQUENCE 785 AA; 85240 MW; 246988E41997DF62 CRC64;

Query Match 36.7%; Score 44; DB 12; Length 785;

Best Local Similarity 41.7%; Pred. No. 15;

Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 12 KQEEEAARLXXXXLKNKGXSSGA 35

Db 422 EQCDEALRRVLRALRGAGGTGA 445

:|:||||:| | | | |

RESULT 3

Q9CIF8 PRELIMINARY; PRT; 310 AA.

AC Q9CIF8; (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE Mevalonate kinase.

GN YEAG OR LL0404.

OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;

OC Streptococcaceae; Lactococcus.

OX NCBI\_TaxID=1360;

RN [1]  
PRELIMINARY; PRT; 546 AA.

RP SEQUENCE FROM N.A.  
RC STRAIN=IL1403;  
RX MEDLINE=21235186; PubMed=11337471;  
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarre K.,  
Weissenbach J., Ehrlich S.D., Sorokin A.;

RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
lactis ssp. lactis IL1403.";

RL Genome Res. 11:731-753(2001).

DR EMBL: AE006277; AAK04502.1; -.

DR InterPro: IPR001745; GHMPkase\_ATP.

DR InterPro: IPR001459; Mew\_gal\_kin.

DR Pfam: PF00288; GHMP\_kinases; 1.

DR PRINTS; PR00959; MEVGALKINASE.

DR Kinase; Complete proteome.

KW SEQUENCE 310 AA; 34334 MW; E85A2C962C9438DA CRC64;

Query Match 35.0%; Score 42; DB 16; Length 310;

Best Local Similarity 33.3%; Pred. No. 13;

Matches 7; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 13 QXEEEAARLXXXXLKNKGXSS 33

Db 285 ENKDAIRISQRLKNGAKNT 305

:|:||||:| | | | |

RESULT 4

Q8ZQW9 PRELIMINARY; PRT; 546 AA.

AC Q8ZQW9; (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 21, Last annotation update)

DE Phosphoglucosyltransferase (EC 5.4.2.2).

GN PGM OR STM0698

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Salmonella.

OX NCBI\_TaxID=602;

RN [1]  
SEQUENCE FROM N.A.

RC STRAIN=LT2 / SGSC1412 / ATCC 700720;

RX MEDLINE=21534948; PubMed=11677609;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,

Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,

Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,

Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,

Waterston R., Wilson R.K.;

RT "Complete genome sequence of Salmonella enterica serovar Typhimurium

LT2.";

RL Nature 413:852-856(2001).

DR EMBL: AE008728; AAL19642.1; -.

DR InterPro: IPR001485; PG/PMM\_mutase.

DR Pfam: PF00408; PGM\_PMM; 1.

DR Pfam: PF02878; PGM\_PMM\_I; 1.

DR Pfam: PF02879; PGM\_PMM\_II; 1.

DR Pfam: PF02880; PGM\_PMM\_III; 1.

DR TIGREAFMS; TIGR01132; pgm; 1.

DR PROSITE; PS00710; PGM\_PMM; 1.

DR Isomerase; Complete proteome.

KW SEQUENCE 546 AA; 58089 MW; A3DD0779F6AE8C95 CRC64;

Query Match 35.0%; Score 42; DB 16; Length 546;

Best Local Similarity 52.9%; Pred. No. 24;

Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 12 KQEEEAARLXXXXLKN 28

Db 529 KQEEEAARLXXXXLKN 545

:|:||||:| | | | |

RESULT 5

Q8Z8F1 PRELIMINARY; PRT; 546 AA.

ID Q8Z8F1



DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Putative deaminase.  
 GN 25CK31.34 OR SC04974.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RN [2]  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Oliver K., Harris D.;  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL451182; CAC18715.2; -;  
 DR EMBL; AL512667; CAD30959.1; -;  
 DR InterPro: IPR02125; dCMP/cyt\_deam.  
 DR Pfam: PF00383; dCMP\_cyt\_deam; 1.  
 DR Hypothetical protein; Complete proteome.  
 SEQUENCE 167 AA; 18334 MW; 3D2044BA11F6E9B1 CRC64;  
 Query Match 34.2%; Score 41; DB 16; Length 167;  
 Best Local Similarity 35.0%; Pred. No. 10;  
 Matches 7; Conservative 5; Mismatches 8; Indels 0; Gaps 0;  
 QY 16 EBAVRLXXXXLKNKGXSSGA 35  
 : : : : :  
 Db 19 DRAIELATTSVRNGGPFGA 38

RESULT 9  
 ID 042143 PRELIMINARY; PRT; 266 AA.  
 AC 042143;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE Glucagon I precursor [contains: Glucagon; glucagon-like peptide 1A  
 DE (GLP-1A); glucagon-like peptide 1B (GLP-1B); glucagon-like peptide 1C  
 DE (GLP-1C); glucagon-like peptide 2 (GLP-2)].  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodinae; Xenopus.

OX NCBI\_TaxID=8355;  
 RN [1]  
 RN [2]  
 RN [3]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC TISSUE=PANCREAS;  
 RX MEDLINE=97368292; PubMed=9223287;  
 RA Irwin D.M., Satkunarajah M., Wen Y., Brubaker P.L., Pederson R.A.,  
 RA Wheeler M.B.;  
 RT "The Xenopus proglucagon gene encodes novel GLP-1-like peptides with  
 RT insulinotropic properties";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:7915-7920(1997).  
 CC -!- FUNCTION: PROMOTES HYDROLYSIS OF GLYCOGEN AND LIPIDS, AND RAISES  
 CC -!- THE BLOOD SUGAR LEVEL.  
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -!- SIMILARITY: BELONGS TO THE GLUCAGON FAMILY.  
 DR EMBL; AF004432; AAB65660.1; -;  
 DR HSSP; P01274; IGCN.  
 DR InterPro: IPR000532; Glucagon.  
 DR Pfam: PF00123; hormone2; 5.  
 DR PRINTS; PR00275; GLUCAGON.  
 DR SMART; SM00070; GLUCA; 5.  
 DR PROSITE; PS00260; GLUCAGON; 5.  
 KW Glucagon family; Hormone; Signal; Cleavage on pair of basic residues;  
 KW Multigene family; Alternative splicing.  
 FT SIGNAL 1 ? POTENTIAL.  
 FT PEPTIDE 53 81  
 FT PEPTIDE 97 133  
 FT PEPTIDE 142 173  
 FT PEPTIDE 180 211  
 FT PEPTIDE 227 259  
 FT VARSPLIC 214 261  
 FT VARSPLIC 214 261  
 SQ SEQUENCE 266 AA; 30951 MW; 544F7BGC20AF872C CRC64;  
 Query Match 34.2%; Score 41; DB 13; Length 266;  
 Best Local Similarity 34.5%; Pred. No. 17;  
 Matches 10; Conservative 5; Mismatches 14; Indels 0; Gaps 0;  
 QY 4 GTXXXXXKXEEEAVALXXXXLKNKGX 32  
 : : : : :  
 Db 100 GTFTSDVTQQLDEKAAKEFDWLINGGPS 128

RESULT 10  
 ID 092527 PRELIMINARY; PRT; 306 AA.  
 AC 092527;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Coat protein (Capsid protein).  
 OS Carnation latent virus (CLV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlaviruses.  
 OX NCBI\_TaxID=12164;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Meenan B.M.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91324119; PubMed=1713905;  
 RA Meenan B.M., Mills P.R.;  
 RT "Nucleotide sequence of the 3'-terminal region of carnation latent  
 RT virus.";  
 RL Intervirology 32:262-267(1991).  
 CC -!- FUNCTION: SELF-ASSEMBLES WITH THE RNA TO FORM INFECTIOUS PARTICLES  
 CC (BY SIMILARITY).  
 CC -!- SIMILARITY: TO THE COAT PROTEINS OF OTHER POTEXVIRUSES.  
 DR EMBL; AJ010697; CAA09306.1; -;  
 DR InterPro: IPR000052; P1v1r\_coat.  
 DR Pfam: PF00286; virus\_P-coat; 1.  
 DR PRINTS; PR00232; POTXCARLCOAT.  
 DR PROSITE; PS000603; P1v1r\_coat; 1.  
 DR PROSITE; PS00418; POTEX\_CARLAVIRUS\_COAT; 1.

```
KW Coat protein.
SQ SEQUENCE 306 AA; 33890 MW; 4456EBB53E174298 CRC64;

Query Match 34.2%; Score 41; DB 12; Length 306;
Best Local Similarity 43.5%; Pred. No. 20;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 12 KXEEEEAVRLXXXXLKNKGXSSG 34
      | | | | | | | | | |
Db 62 KLKEFNSQNLTAGELKNGGFESG 84

RESULT 11
Q9UYT6
ID Q9UYT6 PRELIMINARY; PRT; 402 AA.
AC
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
POLYDNUM cofactor biosynthesis protein (MOEA-1).
PAB1436.
Pyrococcus abyssi.
Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ248287; CAB50326.1; AATRNA_ligaseII.
DR InterPro: IPR002106; MOCF_biosynth.
DR InterPro: IPR001453; MOCF_biosynth.
DR InterPro: IPR005111; MoeA_C.
DR Pfam: PF00994; MOCF_biosynth; 1.
DR Pfam: PF03454; MoeA_C; 1.
DR Pfam: PF03453; MoeA_N; 1.
DR ProDom: PD002460; MOCF_biosynth; 1.
DR TIGRFAMs: TIGR00177; molyb_syn; 1.
DR Complete proteome.
KW PROSITE: PS00339; AA-TRNA_LIGASE-II_2; UNKNOWN_1.
SQ SEQUENCE 402 AA; 43327 MW; 44545EDA70F6A78E CRC64;

Query Match 34.2%; Score 41; DB 17; Length 402;
Best Local Similarity 39.1%; Pred. No. 27;
Matches 9; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

Qy 12 KXEEEEAVRLXXXXLKNKGXSSG 34
      | | | | | | | | | |
Db 237 KELIEGVRVADIVITSGASGG 259

RESULT 12
Q96L69
ID Q96L69 PRELIMINARY; PRT; 589 AA.
AC
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Ectodermal-neural cortex.
EN ENCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Peng Z., Zhang B., Peng X., Yuan J., Qiang B.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY049781; AAL15438.1;
*The genome sequence of Drosophila melanogaster.*;
```

```
DR InterPro: IPR000210; BTB_POZ.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF00651; BTB; 1.
DR Pfam: PF01344; Kelch; 5.
DR PROSITE: PS00097; BTB; 1.
DR PROSITE: PS00639; THIOLEPROTEASE_HIS; UNKNOWN_1.
SQ SEQUENCE 589 AA; 66113 MW; E5CB1466DB8CA16E CRC64;

Query Match 34.2%; Score 41; DB 4; Length 589;
Best Local Similarity 45.0%; Pred. No. 41;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 11 SKXEEEAVRLXXXXLKNKG 30
      | | | | | | | | | |
Db 262 SKEIVEAIRCKLKILQNDG 281

RESULT 13
Q9VRN8
ID Q9VRN8 PRELIMINARY; PRT; 2044 AA.
AC
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Sif protein.
GN SIF OR CG5256 OR CG5406.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananthakrishnan P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B., Stiden-Klamos I., Simpson M., Skupski W.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
*The genome sequence of Drosophila melanogaster.*;
```

```

RL Science 287:2185-2195(2000).
DR EMBL; AE003565; AAF50756.2; -.
DR HSSP; P08567; 1PLS.
DR FlyBase; FBgn0019652; sif.
DR InterPro; IPR001331; GDS_CDC24.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001849; PH.
DR InterPro; IPR003116; RBD.
DR InterPro; IPR000219; RhoGEF.
DR Pfam; PF00169; PH; 2.
DR Pfam; PF02196; RBD; 1.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00233; PH; 2.
DR SMART; SM00455; RBD; 1.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS00741; DH_1; UNKNOWN_1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR SEQUENCE 2044 AA; 228329 MW; 1ACDFBEA63E3FBC1 CRC64;
QY 12 KOXEEAVRLXXXXLKNGXSSGA 35
Db 1761 RQIIRESVRNNSIPMKNGSGS 1784

RESULT 14
Q9VRN7 PRELIMINARY; PRT; 2045 AA.
AC Q9VRN7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Sif protein.
GN SIF OR CG5256 OR CG5406.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,
RA Foster C., Gabriellista A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostali D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.D.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003565; AAF50755.2; -.
DR HSSP; P08567; 1PLS.
DR FlyBase; FBgn0019652; sif.
DR InterPro; IPR001331; GDS_CDC24.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001849; PH.
DR InterPro; IPR003116; RBD.
DR InterPro; IPR000219; RhoGEF.
DR Pfam; PF00169; PH; 2.
DR Pfam; PF02196; RBD; 1.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00233; PH; 2.
DR SMART; SM00455; RBD; 1.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS00741; DH_1; UNKNOWN_1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR SEQUENCE 2045 AA; 228386 MW; A34956429EA3603B CRC64;
QY 12 KOXEEAVRLXXXXLKNGXSSGA 35
Db 1762 RQIIRESVRNNSIPMKNGSGS 1785

Query Match 34.2%; Score 41; DB 5; Length 2044;
Best Local Similarity 41.7%; Pred. No. 1.6e+02;
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 12 KOXEEAVRLXXXXLKNGXSSGA 35
Db 1762 RQIIRESVRNNSIPMKNGSGS 1785

RESULT 15
P96631 PRELIMINARY; PRT; 127 AA.
ID P96631
AC P96631;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Probable repressor protein.
GN YDCN.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RC MEDLINE=98000087; PubMed=9341680;
RX Beloin C., Ayora S., Exley R., Hirschbein L., Ogasawara N.,
RA Kasahara Y., Alonso J.C., Le Hegarat F.;
RT "Characterization of an lrp-like (lrpC) gene from Bacillus subtilis.";
RL Mol. Gen. Genet. 256:63-71(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RC MEDLINE=98044033; PubMed=9384377;
RX Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
```



RA Chai S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
 RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Neone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Serot S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,  
 Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the gram-positive bacterium *Bacillus*  
 subtilis."

RL Nature 390:249-256(1997).  
 RN {3}  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB001488; BAA19320.1; -;  
 DR EMBL; Z99106; CAB12289.1; -;  
 DR InterPro; IPR001387; HTH\_3.  
 DR Pfam; PF01381; HTH\_3; 1.  
 DR SMART; SM00530; HTH\_XRE; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 127 AA; 14649 MW; 3CC91D5B1D51628C CRC64;

Query Match 33.3%; Score 40; DB 16; Length 127;  
 Best Local Similarity 47.1%; Pred. No. 12;  
 Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 13 QXEEAVRLXXXXLKG 29  
 Db : :|||  
 100 EFDEETARLVKKALNG 116

arch completed: January 8, 2003, 11:00:48  
 Job time : 27.1983 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2003, 10:54:49 ; Search time 51.1034 Seconds  
(without alignments)  
99.084 Million cell updates/sec

Title: US-09-003-869-4

Perfect score: 46  
Sequence: 1 XXXGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 38

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

al number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID        | Description        |
|------------|-------|-------------|--------|-----------|--------------------|
| 1          | 10    | 21.7        | 96     | ABG16827  | Novel human diagno |
| 2          | 10    | 21.7        | 150    | ABG61831  | Prostate cancer-as |
| 3          | 10    | 21.7        | 218    | AAAG00755 | Human secreted pro |
| 4          | 10    | 21.7        | 218    | AAAG00759 | Human secreted pro |
| 5          | 10    | 21.7        | 220    | AAU76972  | Human ribosomal L1 |
| 6          | 10    | 21.7        | 222    | AAU79212  | Corynebacterium gl |
| 7          | 10    | 21.7        | 225    | AAAG00758 | Human secreted pro |
| 8          | 10    | 21.7        | 226    | AAW05151  | Nuclear proliferat |
| 9          | 10    | 21.7        | 265    | AAAB58221 | Lung cancer associ |
| 10         | 10    | 21.7        | 285    | ABB05433  | Rat BNP receptor   |

|    |    |      |      |    |          |                    |
|----|----|------|------|----|----------|--------------------|
| 11 | 10 | 21.7 | 314  | 18 | AAW14283 | Human neuroblastom |
| 12 | 10 | 21.7 | 314  | 21 | AAV70572 | Mouse Phox2b prote |
| 13 | 10 | 21.7 | 327  | 22 | ABB66932 | Drosophila melanog |
| 14 | 10 | 21.7 | 330  | 22 | ACG90013 | C glutamicum prote |
| 15 | 10 | 21.7 | 336  | 21 | ACG16755 | Arabidopsis thalia |
| 16 | 10 | 21.7 | 367  | 23 | ABB10101 | Human homeobox pro |
| 17 | 10 | 21.7 | 388  | 23 | ABG63243 | Human prostate spe |
| 18 | 10 | 21.7 | 388  | 23 | ABG61877 | Drosophila melanog |
| 19 | 10 | 21.7 | 431  | 22 | ABG69274 | Drosophila melanog |
| 20 | 10 | 21.7 | 443  | 22 | AAE05813 | Human small cell l |
| 21 | 10 | 21.7 | 462  | 22 | ABG61682 | Drosophila melanog |
| 22 | 10 | 21.7 | 462  | 22 | AAU38933 | Drosophila G-prote |
| 23 | 10 | 21.7 | 481  | 23 | ABG92770 | Human protein sequ |
| 24 | 10 | 21.7 | 481  | 23 | ABG97545 | Novel human protei |
| 25 | 10 | 21.7 | 526  | 22 | ABG62199 | Drosophila melanog |
| 26 | 10 | 21.7 | 536  | 22 | ABG60107 | Drosophila melanog |
| 27 | 10 | 21.7 | 644  | 22 | ABG59973 | Drosophila melanog |
| 28 | 10 | 21.7 | 645  | 22 | ABG61545 | Drosophila melanog |
| 29 | 10 | 21.7 | 675  | 22 | ABG71143 | Drosophila melanog |
| 30 | 10 | 21.7 | 738  | 19 | AAW56163 | New DNA sequence i |
| 31 | 10 | 21.7 | 791  | 22 | ABG23551 | Novel human diagno |
| 32 | 10 | 21.7 | 823  | 22 | ABG61552 | Drosophila melanog |
| 33 | 10 | 21.7 | 852  | 22 | ABG58079 | Drosophila melanog |
| 34 | 10 | 21.7 | 871  | 22 | ABG58282 | Drosophila melanog |
| 35 | 10 | 21.7 | 872  | 22 | ABG70960 | Drosophila melanog |
| 36 | 10 | 21.7 | 883  | 22 | ABG62230 | Drosophila melanog |
| 37 | 10 | 21.7 | 998  | 22 | ABG60423 | Drosophila melanog |
| 38 | 10 | 21.7 | 1182 | 22 | ABG64438 | Drosophila melanog |
| 39 | 10 | 21.7 | 1186 | 22 | ABG63516 | Drosophila melanog |
| 40 | 10 | 21.7 | 1237 | 21 | AAV81609 | Streptococcus pneu |
| 41 | 10 | 21.7 | 1380 | 20 | AAV08402 | Drosophila sp. ROB |
| 42 | 10 | 21.7 | 1381 | 20 | AAV13564 | Drosophila Robo 2  |
| 43 | 10 | 21.7 | 1463 | 23 | AAE20110 | Lactobacillus rham |
| 44 | 10 | 21.7 | 1468 | 22 | ABG65329 | Drosophila melanog |
| 45 | 10 | 21.7 | 1518 | 22 | ABG64829 | Drosophila melanog |

#### ALIGNMENTS

RESULT 1  
ABG16827  
ID ABG16827 standard; Protein: 96 AA.  
XX  
AC ABG16827;  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #16818.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI: 2001-639362/73.  
XX N-PSDB; AA581014.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
PT

PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX Claim 20; SEQ ID NO 47186; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful for treating  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG0010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 96 AA;

Query Match 21.7%; Score 10; DB 22; Length 96;  
Best Local Similarity 13.3%; Pred. No. 84;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXA 18  
|  
Db 2 GTAAAAAATAAAAAA 16

RESULT 2

ABG61831  
ID ABG61831 standard; Protein; 150 AA.

XX AC ABG61831;

XX DT 15-AUG-2002 (first entry)

XX DE Prostate cancer-associated protein #32.

XX Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.

XX OS Mammalia.

XX PN WO200230268-A2.

XX PD 18-APR-2002.

XX PF 12-OCT-2001; 2001WO-US32045.

XX PR 13-OCT-2000; 2000US-0687576.

XX PR 08-DEC-2000; 2000US-0733288.

XX PR 08-DEC-2000; 2000US-0733742.

XX PR 24-JAN-2001; 2001US-263957P.

XX PR 16-MAR-2001; 2001US-276791P.

XX PR 16-MAR-2001; 2001US-276888P.

XX PR 06-APR-2001; 2001US-281922P.

XX PR 24-APR-2001; 2001US-286214P.

XX PR 30-APR-2001; 2001US-0847046.

XX PR 04-MAY-2001; 2001US-288589P.

XX (EOSB-) EOS BIOTECHNOLOGY INC.

XX Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;

XX

DR WPI; 2002-471335/50.  
DR N-PSDB; ABK92146.

XX Detecting a prostate cancer-associated transcript in a cell in a  
PT patient, useful for diagnosing prostate cancer (PC) or screening  
PT modulators of PC, by determining if prostate cancer-associated genes  
PT are expressed in a prostate tissue.

XX Claim 27; Page 326; 436pp; English.

XX The present invention relates to methods of detecting a prostate  
CC cancer-associated transcript in a cell from a patient. The method  
CC comprises contacting a biological sample from the patient with  
CC prostate cancer-associated polynucleotides (designated PC genes) that  
CC selectively hybridise to a sequence that is at least 80% identical  
CC to them. The prostate cancer-associated polynucleotide sequences  
CC are differentially expressed in prostate tumour tissue or in  
CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).  
CC The methods of the invention are useful for diagnosing and treating  
CC prostate cancer in mammals. The prostate cancer-associated genes are  
CC useful for diagnosing or treating prostate cancer, as well as for  
CC identifying modulators of prostate cancer or agents that inhibit  
CC prostate cancer. The nucleic acid sequences are particularly useful  
CC in gene therapy, as a vaccine or in antisense applications.  
CC ABG61800-ABG61944 represent prostate cancer-associated proteins.

XX Sequence 150 AA;

Query Match 21.7%; Score 10; DB 23; Length 150;  
Best Local Similarity 13.3%; Pred. No. 96;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXA 18  
|  
Db 126 GTTAAAAAATAAAAAA 140

RESULT 3

AAG00755

ID AAG00755 standard; Protein; 218 AA.

XX AC AAG00755;

XX DT 06-OCT-2000 (first entry)

XX DE Human secreted protein, SEQ ID NO: 4836.

XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping.

XX OS Homo sapiens.

XX PN EP1033401-A2.

XX PD 06-SEP-2000.

XX PF 21-FEB-2000; 2000EP-0200610.

XX PR 26-FEB-1999; 99US-0122487.

XX PA (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.

DR N-PSDB; AAC00761.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 13; SEQ ID 4836; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number  
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
 CC different tissues. EST sequences usually correspond mainly to the 3'  
 CC untranslated region (UTR) of the mRNA because they are often obtained  
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
 CC those cases where longer cDNA sequences have been obtained, the full 5'  
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
 CC ends and can therefore be used to obtain full length cDNAs and genomic  
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
 CC chromosome mapping procedures. They are used to obtain upstream  
 CC regulatory sequences and to design expression and secretion vectors.

XX Sequence 218 AA;

Query Match 21.7%; Score 10; DB 21; Length 218;  
 Best Local Similarity 13.3%; Pred. No. 1.1e+02;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

4 GXXXXXXXXXXXXXA 18  
 |  
 148 GTAAAAAATAAAAAA 162

DB  
 RESULT 4  
 AAG00759  
 ID AAG00759 standard; Protein; 218 AA.  
 AC AAG00759;  
 XX  
 XX 06-OCT-2000 (first entry)  
 DT  
 DE Human secreted protein, SEQ ID NO: 4840.  
 XX  
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 (GEST ) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI; 2000-500381/45.  
 DR N-PSDB; AAC00765.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 PS Claim 13; SEQ ID 4840; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number  
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
 CC different tissues. EST sequences usually correspond mainly to the 3'  
 CC untranslated region (UTR) of the mRNA because they are often obtained  
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
 CC those cases where longer cDNA sequences have been obtained, the full 5'  
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
 CC ends and can therefore be used to obtain full length cDNAs and genomic  
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and

CC chromosome mapping procedures. They are used to obtain upstream  
 CC regulatory sequences and to design expression and secretion vectors.  
 XX  
 SQ Sequence 218 AA;

Query Match 21.7%; Score 10; DB 21; Length 218;  
 Best Local Similarity 13.3%; Pred. No. 1.1e+02;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 QY 4 GXXXXXXXXXXXXXA 18  
 |  
 DB 148 GTAAAAAATAAAAAA 162

RESULT 5

AAU76972  
 ID AAU76972 standard; Protein; 220 AA.

XX  
 AC AAU76972;

XX 21-MAY-2002 (first entry)

XX Human ribosomal L14 (RPL14) protein.

XX Ribosomal L14 protein; RPL14; CD39L3; PMGM; GC20; cancer; metastasis;  
 KW carcinoma; non-small cell carcinoma; smoking; lung cancer;  
 KW bladder cancer; head cancer; neck cancer; urothelial cancer;  
 KW kidney cancer; pancreas cancer; mouth cancer; throat cancer;  
 KW pharynx cancer; larynx cancer; upper airway primary cancer;  
 KW upper airway secondary cancer; esophagus cancer; chromosome 3p21.3.

XX Homo sapiens.

XX WO200212563-A2.

XX 14-FEB-2002.

XX 06-AUG-2001; 2001WO-US24718.

XX 04-AUG-2000; 2000US-222811P.

XX (TEXA ) UNIV TEXAS SYSTEM.

XX Katz R, Jiang F;

XX WPI; 2002-217200/27.

XX N-PSDB; ABK10349.

PT Identifying subject at risk for development of cancer, preferably lung  
 PT cancer, comprises contacting RPL14, CD39L3, PMGM, or GC20 gene probe  
 PT with test sample obtained from subject, and analysing DNA from test  
 PT sample -

XX Example 1; Page 69-70; 79pp; English.

XX The invention describes a method of identifying a subject at risk for the  
 CC development of cancer, predicting progression or metastasis of non-small  
 CC cell carcinoma and other carcinoma in a subject, or identifying an  
 CC individual to be segregated from a high risk environment. The method  
 CC comprises contacting an RPL14, CD39L3, PMGM, or GC20 gene probe with a  
 CC test sample obtained from a subject, and analysing DNA from the test  
 CC sample. The method is useful for identifying a subject (a smoker,  
 CC non-smoker or former smoker) at risk for the development, recurrence, or  
 CC metastasis of cancer (preferably cancer of lung, bladder, head, neck,  
 CC urothelial, kidney, pancreas, mouth, throat, pharynx, larynx or  
 CC esophagus, or an upper airway primary or secondary cancer), to identify  
 CC subjects who need an intensive follow-up protocol and for the prognosis  
 CC and diagnosis of cancer. This is the amino acid sequence of the human  
 CC ribosomal L14 protein (RPL14, located on chromosome 3p21.3), used to  
 CC develop the gene probe described in the method of the invention.

XX Sequence 220 AA;

Query Match 21.7%; Score 10; DB 23; Length 220;  
 Best Local Similarity 13.3%; Pred. No. 1.1e+02;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXX 18  
 DB 148 GTAAAAA 162

RESULT 6  
 ID AAB79212 standard; Protein; 222 AA.  
 AC AAB79212;  
 DT 30-APR-2001 (first entry)  
 XX Corynebacterium glutamicum HA protein sequence SEQ ID NO:380.

Corynebacterium glutamicum; homeostasis; adaptation; HA protein;  
 fine chemical production; organic acid; proteinogenic amino acid;  
 nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;  
 diagnosis; Corynebacterium diphtheriae; genetic engineering;  
 Brevibacterium; environmental condition.

XX Corynebacterium glutamicum.  
 OS WO200100842-A2.  
 XX 04-JAN-2001.  
 XX 23-JUN-2000; 2000WO-IB00911.  
 XX 25-JUN-1999; 99US-0141031.  
 PR 08-JUL-1999; 99DE-1031636.  
 PR 09-JUL-1999; 99DE-1032125.  
 PR 09-JUL-1999; 99DE-1032126.  
 PR 09-JUL-1999; 99DE-1032127.  
 PR 09-JUL-1999; 99DE-1032128.  
 PR 09-JUL-1999; 99DE-1032129.  
 PR 09-JUL-1999; 99DE-1032226.  
 PR 14-JUL-1999; 99DE-1032320.  
 PR 14-JUL-1999; 99DE-1032922.  
 PR 14-JUL-1999; 99DE-1032924.  
 PR 14-JUL-1999; 99DE-1032928.  
 PR 14-JUL-1999; 99DE-1032930.  
 PR 14-JUL-1999; 99DE-1032933.  
 PR 14-JUL-1999; 99DE-1032935.  
 PR 14-JUL-1999; 99DE-1032973.  
 PR 14-JUL-1999; 99DE-1033002.  
 PR 14-JUL-1999; 99DE-1033003.  
 PR 14-JUL-1999; 99DE-1033005.  
 PR 14-JUL-1999; 99DE-1033006.  
 PR 31-AUG-1999; 99DE-1041378.  
 PR 31-AUG-1999; 99DE-1041379.  
 PR 31-AUG-1999; 99DE-1041390.  
 PR 31-AUG-1999; 99DE-1041391.  
 PR 03-SEP-1999; 99DE-1042088.  
 XX (BADI ) BASF AG.  
 XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
 XX WPI: 2001-061974/07.  
 DR N-PSDB; AAB71327.  
 DR

PS Claim 20; Page 633-634; 712pp; English.  
 CC AAB71138 to AAB71357 encode the Corynebacterium glutamicum homeostasis  
 CC and adaptation (HA) proteins given in AAB79212 to AAB79242. The  
 CC C. glutamicum HA genes (I) can be used in vectors for expression in host  
 CC cells and production of fine chemicals, such as, an organic acid,  
 CC proteinogenic or nonproteinogenic amino acid (preferred), purine or  
 CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated  
 CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,  
 CC polyketide or enzyme. The amino acids produced can be lysine, glutamine,  
 CC glutamate, alanine, aspartate, glycine, serine, threonine, methionine,  
 CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,  
 CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can  
 CC be modulated. The presence of (I) or HA proteins encoded by then are  
 CC used for diagnosing the presence or activity of Corynebacterium  
 CC diphtheriae. (I) can be used to map the C. glutamicum genome or can be  
 CC used as markers for genetically engineered Corynebacterium or  
 CC Brevibacterium. The HA proteins encoded by the (I) are used to maintain  
 CC homeostasis in C. glutamicum or help the microorganism to adapt to  
 CC different environmental conditions.

XX Sequence 222 AA;  
 SQ Query Match 21.7%; Score 10; DB 22; Length 222;  
 Best Local Similarity 13.3%; Pred. No. 1.1e+02;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXX 18  
 DB 58 GSSAASSTSSASSA 72

RESULT 7  
 ID AAG00758 standard; Protein; 225 AA.  
 AC AAG00758;  
 XX 06-OCT-2000 (first entry)  
 DT Human secreted protein, SEQ ID NO: 4839.  
 DE Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping.  
 KW Homo sapiens.  
 OS EP1033401-A2.  
 XX PD 06-SEP-2000.  
 XX PF 21-FEB-2000; 2000EP-0200610.  
 XX PR 26-FEB-1999; 99US-0122487.  
 XX (GEST ) GENSET.  
 PA Dumas Milne Edwards J, Duclert A, Giordano J;  
 PI WPI: 2000-500381/45.  
 XX DR N-PSDB; AAC00764.  
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX Claim 13; SEQ ID 4839; 71pp + CD-ROM; English.  
 PS The present sequence is a polypeptide encoded by one of a large number  
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
 CC different tissues. EST sequences usually correspond mainly to the 3'  
 CC untranslated region (UTR) of the mRNA because they are often obtained

CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
 CC those cases where longer cDNA sequences have been obtained, the full 5'  
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
 CC ends and can therefore be used to obtain full length cDNAs and genomic  
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
 CC chromosome mapping procedures. They are used to obtain upstream  
 CC regulatory sequences and to design expression and secretion vectors.

XX Sequence 225 AA;

Query Match 21.7%; Score 10; DB 21; Length 225;  
 Best Local Similarity 13.3%; Pred. No. 1.1e+02;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXXA 18

DB 148 GTAAAAAATAAAAAA 162

ULT 8  
 05151

ID AAN05151 standard; peptide; 226 AA.

AC AAN05151;

DT 17-JUN-1997 (first entry)

DE Nuclear proliferation-associated antigen p28.

KW Human cell proliferation; tumour; antigen; mammary carcinoma;  
 diagnosis; prognosis; proliferative activity; malignant lymphoma;  
 endometrial cycle status; antigen-stimulated lymphocyte proliferation.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Region 3..226

FT /note= "This region of the sequence is claimed,  
 1.e. it corresponds to SEQ.ID.No.3"

FT Misc-difference 109

FT /note= "undefined"

PN EP736543-A2.

XX 09-OCT-1996.

XX 03-APR-1996; 96EP-0105373.

XX 04-APR-1995; 95DE-4012504.

PA (BOE ) BOEHRINGER MANNHEIM GMBH.

XX Parwaresch R;

XX WPI; 1996-444847/45.

XX Nuclear proliferation-associated antigens and monoclonal antibodies  
 PT against them - useful for diagnosis of proliferative state of a cell

XX Claim 30; Fig 5; 36pp; German.

XX Antigens that have molecular weights of ca. 24 and 28 kD by SDS-PAGE  
 CC and are obtained from the nuclei of human proliferating cells by  
 CC immunoreaction with monoclonal antibody Anti-p28/24 produced by  
 CC hybridoma cell line DSM ACC 2199 are claimed. The present sequence  
 CC is from the p28 antigen. Antibodies specific for the nuclear  
 CC proliferation-associated p24 and p28 antigens do not cross-react with  
 CC cytoplasmic material or with nuclei from non-proliferating cells.  
 CC Such antibodies, their fragments and conjugates are useful as markers  
 CC for determining the proliferative activity of cells, especially as ELISA  
 CC reagents for prognosis of mammary carcinoma or malignant lymphoma, for  
 CC assessing endometrial cycle status, for distinguishing between benign

CC and malignant melanomas and for determination of antigen-stimulated  
 CC lymphocyte proliferation.

XX Sequence 226 AA;

Query Match 21.7%; Score 10; DB 17; Length 226;  
 Best Local Similarity 13.3%; Pred. No. 1.1e+02;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXXA 18

DB 149 GTAAAAAATAAAAAA 163

RESULT 9

AAB58221

ID AAB58221 standard; Protein; 265 AA.

XX AAB58221;

DT 14-MAR-2001 (first entry)

DE Lung cancer associated polypeptide sequence SEQ ID 559.

XX Human; lung cancer associated protein; neuroprotective; cytostatic;  
 KW cardioactive; immunomodulatory; muscular active; vulnerary;  
 KW gastrointestinal; nephrotropic; antiinfective; gynecological;  
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;  
 KW proliferative disorder; wound healing; infectious disease.

OS Homo sapiens.

XX WO200055180-A2.

PN 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05918.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (ROSE/) ROSEN C A.

XX Ruben SM;

XX WPI; 2000-587514/55.

XX N-PSDB; AAF18097.

XX Lung cancer associated gene sequences, referred to as lung cancer  
 PT antigens, useful for treatment, prevention, and diagnosis of disorders  
 PT such as lung cancer -

XX Claim 11; Page 1054-1055; 1425pp; English.

XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer  
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer  
 CC associated proteins and polynucleotide sequences, their agonists, and  
 CC antagonists may have neuroprotective; cytostatic; cardioactive;  
 CC immunomodulatory; muscular active general; vulnerary; gastrointestinal  
 CC general; nephrotropic; antiinfective; gynecological; or antibacterial  
 CC activity. The invention also includes antibodies specific for the  
 CC protein or polynucleotide sequences. The lung cancer associated  
 CC polynucleotide sequences may be used for detection of lung cancer,  
 CC chromosome identification, as chromosome markers, and for numerous other  
 CC diagnostic or research purposes. The proteins may be used to treat  
 CC disorders such as neural, immune, muscular, reproductive,  
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
 CC disorders. The proteins may also be used in the treatment of wounds and  
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and  
 CC peptide AAB58549 are used in the course of the invention for the  
 CC identification and characterisation of the polynucleotide and protein  
 CC sequences.

XX

SQ Sequence 265 AA;

Query Match 21.7%; Score 10; DB 21; Length 265;  
Best Local Similarity 13.3%; Pred. No. 1.1e+02;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXX 18  
|  
DB 193 GTAAAAA 207

RESULT 10

AB05433  
ID ABB05433 standard; Protein; 285 AA.

XX AC ABB05433;

XX DT 15-APR-2002 (first entry)

XX Rat BINP receptor function related protein SEQ ID NO:2.

KW Rat; anti-BINP-binding protein monoclonal antibody 6A22; neuroprotective;  
KW brain injury-derived neurotrophic peptide; BINP receptor function;  
KW gene therapy; nerve function activator.

XX OS Rattus norvegicus.

XX PN JP2001321189-A.

XX PD 20-NOV-2001.

XX PF 08-MAR-2001; 2001JP-0064883.

XX PR 10-MAR-2000; 2000JP-0065895.

XX PA (MITU ) MITSUBISHI CHEM CORP.

DR WPI; 2002-144135/19.

DR N-PSDB; ABA93115.

XX New DNA encoding BINP receptor protein, useful for screening for BINP  
PT receptor agonists as nerve function activators .

XX Claim 4; Page 13-14; 15pp; Japanese.

XX The present sequence represents a rat protein which has brain injury-  
CC derived neurotrophic peptide (BINP) receptor function (I). (I) has  
CC neuroprotective activity. The polynucleotide sequence (II) encoding (I)  
CC can be used in gene therapy. The BINP receptor protein can be used for  
CC screening for BINP receptor agonists. The agonists are useful as nerve  
CC function activators. The present invention also describes the  
CC anti-BINP-binding protein monoclonal antibody 6A22.

XX SQ Sequence 285 AA;

Query Match 21.7%; Score 10; DB 23; Length 285;  
Best Local Similarity 13.3%; Pred. No. 1.2e+02;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXX 18  
|  
DB 260 GTTAAAAA 274

RESULT 11

AAW14283

ID AAW14283 standard; Protein; 314 AA.

XX AC AAW14283;

XX DT 09-JAN-1998 (first entry)

XX Human neuroblastoma-specific paired-like homeobox protein sequence.

XX

KW Human; neuroblastoma; paired; homeobox; glioma; probe; diagnosis;  
XX detection; tumour.

XX OS Homo sapiens.

XX PN JP09191883-A.

XX PD 29-JUL-1997.

XX PF 16-JAN-1996; 96JP-0004729.

XX PR 16-JAN-1996; 96JP-0004729.

XX PA (NISB ) JAPAN TOBACCO INC.

XX DR WPI; 1997-429183/40.

XX DR N-PSDB; AAT85971.

XX DNA encoding human neuroblastoma-specific paired-like homeobox  
PT protein - useful for detecting glioma(s) and tumours

XX Claim 1; Page 10-11; 14pp; Japanese.

XX This is the amino acid sequence of a novel neuroblastoma-specific  
CC paired-like homeobox protein which is expressed on human glioma cells.  
CC The encoding gene was isolated from a 3' directed cDNA library prepared  
CC from human neuroblastoma cell line CHPL34. The screen isolated a clone  
CC designated GS0886 whose insert contained the coding sequence and the  
CC 5' and 3' sequences of the gene (AAT85972-3 respectively). Expression  
CC of the gene was detected in neuroblastoma cell lines. Oligonucleotides  
CC derived from the sequence of the homeobox gene can be used as probes  
CC for diagnosing human gliomas, and in the detection of new tumours.

XX SQ Sequence 314 AA;

Query Match 21.7%; Score 10; DB 18; Length 314;  
Best Local Similarity 13.3%; Pred. No. 1.2e+02;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXX 18  
|  
DB 240 GAAAAA 254

RESULT 12

AA70572  
ID AAY70572 standard; Protein; 314 AA.

XX AC AAY70572;

XX DT 04-JUL-2000 (first entry)

XX DE Mouse Phox2b protein.

XX Neurogenin; non-neuronal cell; NNC; neurogenesis; tyrosine hydroxylase;  
KW Phox2b protein; neuronal subtype-specific marker; growth factor;  
KW neural differentiation; transplantation; neuronal dysfunction;  
KW optical nerve damage; auditory nerve damage; neurodegenerative disorder;  
KW neuroprotective; nootropic; anticonvulsant; antiparkinsonian; vulnary;  
KW cerebrotective; immunosuppressant; antiinfectious.

XX OS Mus sp.

XX PN WO200009676-A2.

XX PD 24-FEB-2000.

XX PF 13-AUG-1999; 99WO-US18525.

XX PR 14-AUG-1998; 98US-0096630.

XX PA (CALY ) CALIFORNIA INST OF TECHNOLOGY.



XX Anderson DJ, Lo L;  
 PI WPI: 2000-256250/22.  
 DR N-PSDB; AA251983.  
 XX  
 CC Inducing non-neuronal cells to differentiate into neurons and for  
 CC non-neuronal cells to express a neuronal subtype-specific marker,  
 PT comprising contacting the non-neuronal cells with a vector containing  
 PT neurogenin nucleic acid -  
 XX  
 PS Disclosure; Fig 2D; 76pp; English.  
 XX  
 CC The patent discloses a method for inducing non-neuronal cells (NNC) to  
 CC differentiate into neurons and for NNCs to express a neuronal subtype  
 CC -specific marker. Transformed host cells are used as sources of neuronal  
 CC and other growth factors; in culture for screening compounds that  
 CC modulate neural differentiation or as sources of recombinantly produced  
 CC neurogenins and Phox2a proteins for use in transplantation. The cells  
 CC also have a variety of in vivo uses, e.g. for transplantation at sites of  
 CC neuronal dysfunction e.g. patients with hearing or vision loss due to  
 CC optical or auditory nerve damage, brain or spinal cord injuries, and  
 CC neurodegenerative disorders e.g. Alzheimer's disease. NNCs differentiate  
 CC into neurons through the recombinant expression of a transcription factor  
 CC that induces a core program of neurogenesis. The present sequence encodes  
 CC mouse Phox2b protein. Forced expression of mouse Phox2b can produce  
 CC neurons which express tyrosine hydroxylase (TH), the rate limiting enzyme  
 CC in the synthesis of catecholamines.  
 XX  
 CC Sequence 314 AA;  
 SQ  
 Query Match 21.7%; Score 10; DB 21; Length 314;  
 Best Local Similarity 13.3%; Pred. No. 1.2e+02;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 Qy 4 GXXXXXXXXXXXXXA 18  
 | |  
 Db 240 GAAAAAIAAAAAAAAAA 254  
 RESULT 13  
 ID ABB66932  
 XX ABB66932 standard; Protein; 327 AA.  
 AC ABB66932;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 XX Drosophila melanogaster polypeptide SEQ ID NO 27588.  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX Drosophila melanogaster.  
 OS  
 XX WO200171042-A2.  
 PN  
 XX 27-SEP-2001.  
 PD  
 XX 23-MAR-2001; 2001WO-US09231.  
 PF  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX (PEKE ) PE CORP NY.  
 PA  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI WPI: 2001-656860/75.  
 DR N-PSDB; ABL11035.  
 XX  
 CC New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -  
 XX Disclosure; SEQ ID NO 27588; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 327 AA;  
 Query Match 21.7%; Score 10; DB 22; Length 327;  
 Best Local Similarity 13.3%; Pred. No. 1.2e+02;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 Qy 4 GXXXXXXXXXXXXXA 18  
 | |  
 Db 213 GASTSAAAAAIAAAAAA 227  
 RESULT 14  
 ID AAG90013  
 XX AAG90013 standard; Protein; 330 AA.  
 AC AAG90013;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 XX C glutamicum protein fragment SEQ ID NO: 3767.  
 DE Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 KW Corynebacterium glutamicum.  
 OS  
 XX EP1108790-A2.  
 PN  
 XX 20-JUN-2001.  
 PD  
 XX 18-DEC-2000; 2000EP-0127688.  
 PF  
 XX 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA  
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX WPI: 2001-376931/40.  
 DR N-PSDB; AAH65232.  
 XX  
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 XX Claim 17; SEQ ID NO: 3767; 246pp + Sequence Listing; English.  
 PS  
 XX The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived

CC from corynebacterium. Corynebacterium bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.

XX Sequence 330 AA;

SQ Query Match 21.7%; Score 10; DB 22; Length 330;

Best Local Similarity 13.3%; Pred. No. 1.2e+02;

Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18

DB 58 GSSAASSTSSASSA 72

JLT 15

16755

AAG16755 standard; Protein; 336 AA.

XX AAG16755;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 17522.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 23-APR-1999; 99US-0130891.

XX 28-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 30-APR-1999; 99US-0132407.

XX 04-MAY-1999; 99US-0132484.

XX 05-MAY-1999; 99US-0132489.

XX 06-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132487.

XX 11-MAY-1999; 99US-0132863.

XX 14-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 14-MAY-1999; 99US-0134370.

XX 18-MAY-1999; 99US-0134768.

XX 19-MAY-1999; 99US-0134941.

XX 20-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 06-AUG-1999; 99US-0147192.  
PR 07-AUG-1999; 99US-0147260.  
PR 08-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 10-AUG-1999; 99US-0147493.  
PR 11-AUG-1999; 99US-0147935.  
PR 12-AUG-1999; 99US-0148171.  
PR 13-AUG-1999; 99US-0148319.  
PR 14-AUG-1999; 99US-0148341.  
PR 15-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149368.  
PR 18-AUG-1999; 99US-0149175.  
PR 19-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 21-AUG-1999; 99US-0149723.  
PR 22-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 24-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 28-AUG-1999; 99US-0151066.  
PR 29-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 02-SEP-1999; 99US-0152363.  
PR 03-SEP-1999; 99US-0153070.  
PR 04-SEP-1999; 99US-0153758.  
PR 05-SEP-1999; 99US-0154018.  
PR 06-SEP-1999; 99US-0154039.  
PR 07-SEP-1999; 99US-0154779.  
PR 08-SEP-1999; 99US-0155139.  
PR 09-SEP-1999; 99US-0155486.  
PR 10-SEP-1999; 99US-0155659.  
PR 11-SEP-1999; 99US-0156458.  
PR 12-SEP-1999; 99US-0156596.  
PR 13-SEP-1999; 99US-0157117.  
PR 14-SEP-1999; 99US-0157753.  
PR 15-SEP-1999; 99US-0157865.  
PR 16-SEP-1999; 99US-0158029.  
PR 17-SEP-1999; 99US-0158232.  
PR 18-SEP-1999; 99US-0158369.  
PR 19-SEP-1999; 99US-0159293.  
PR 20-SEP-1999; 99US-0159294.  
PR 21-SEP-1999; 99US-0159295.  
PR 22-SEP-1999; 99US-0159329.  
PR 23-SEP-1999; 99US-0159330.  
PR 24-SEP-1999; 99US-0159331.  
PR 25-SEP-1999; 99US-0159637.  
PR 26-SEP-1999; 99US-0159638.  
PR 27-SEP-1999; 99US-0159584.  
PR 28-SEP-1999; 99US-0160741.  
PR 29-SEP-1999; 99US-0160767.  
PR 30-SEP-1999; 99US-0160768.  
PR 01-OCT-1999; 99US-0160770.  
PR 02-OCT-1999; 99US-0160814.  
PR 03-OCT-1999; 99US-0160815.  
PR 04-OCT-1999; 99US-0160980.  
PR 05-OCT-1999; 99US-0160981.  
PR 06-OCT-1999; 99US-0160989.  
PR 07-OCT-1999; 99US-0161404.  
PR 08-OCT-1999; 99US-0161405.  
PR 09-OCT-1999; 99US-0161406.  
PR 10-OCT-1999; 99US-0161359.  
PR 11-OCT-1999; 99US-0161360.  
PR 12-OCT-1999; 99US-0161361.  
PR 13-OCT-1999; 99US-0161320.  
PR 14-OCT-1999; 99US-0161992.

PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
Query Match 21.7%; Score 10; DB 21; Length 336;  
Best Local Similarity 13.3%; Pred. No. 1.2e+02;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
Qy 4 GXXXXXXXXXXXXXA 18  
Db 5 GATSTSAAXXXXXX 19

Search completed: January 8, 2003, 10:58:23  
Job time : 53.1034 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2003, 10:54:49 ; Search time 10.8103 Seconds  
(without alignments)  
103.426 Million cell updates/sec

Title: US-09-003-869-4  
Perfect score: 46  
Sequence: 1 XXGXGXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 38

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A-COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B-COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A-COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B-COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS-COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description       |
|------------|-------|-------------|--------|-------|-------------------|
| 1          | 10    | 21.7        | 223    | 4     | US-09-009-816-4   |
| 2          | 10    | 21.7        | 367    | 4     | US-09-009-816-2   |
| 3          | 10    | 21.7        | 738    | 3     | US-08-864-038A-3  |
| 4          | 10    | 21.7        | 1345   | 2     | US-08-977-767-3   |
| 5          | 10    | 21.7        | 1381   | 4     | US-09-540-245A-16 |
| 6          | 10    | 21.7        | 1652   | 4     | US-09-627-650B-1  |
| 7          | 10    | 21.7        | 1652   | 4     | US-09-436-063C-1  |
| 8          | 9     | 19.6        | 182    | 1     | US-08-240-712-29  |
| 9          | 9     | 19.6        | 182    | 1     | US-08-443-890-29  |
| 10         | 9     | 19.6        | 228    | 2     | US-08-997-080-147 |
| 11         | 9     | 19.6        | 228    | 2     | US-08-997-362-147 |
| 12         | 9     | 19.6        | 228    | 4     | US-09-095-855-147 |
| 13         | 9     | 19.6        | 228    | 4     | US-09-324-542-147 |
| 14         | 9     | 19.6        | 228    | 4     | US-09-205-426-147 |
| 15         | 9     | 19.6        | 277    | 4     | US-09-186-276B-34 |
| 16         | 9     | 19.6        | 277    | 4     | US-08-842-445-34  |
| 17         | 9     | 19.6        | 277    | 4     | US-09-186-188B-34 |
| 18         | 9     | 19.6        | 340    | 5     | PCT-US96-02331-13 |
| 19         | 9     | 19.6        | 548    | 2     | US-08-871-266B-17 |
| 20         | 9     | 19.6        | 548    | 2     | US-09-018-864A-17 |
| 21         | 9     | 19.6        | 548    | 3     | US-08-871-267B-23 |
| 22         | 9     | 19.6        | 548    | 4     | US-09-618-419-23  |
| 23         | 9     | 19.6        | 842    | 5     | PCT-US96-02331-15 |
| 24         | 9     | 19.6        | 1128   | 4     | US-09-627-650B-11 |
| 25         | 9     | 19.6        | 1128   | 4     | US-09-436-063C-11 |
| 26         | 9     | 19.6        | 1312   | 4     | US-09-268-866-2   |
| 27         | 9     | 19.6        | 2508   | 4     | US-09-627-650B-7  |

|    |   |      |      |   |                  |                    |
|----|---|------|------|---|------------------|--------------------|
| 28 | 9 | 19.6 | 2508 | 4 | US-09-436-063C-7 | Sequence 7, Appli  |
| 29 | 9 | 19.6 | 2544 | 4 | US-09-627-650B-3 | Sequence 3, Appli  |
| 30 | 9 | 19.6 | 2544 | 4 | US-09-436-063C-3 | Sequence 3, Appli  |
| 31 | 9 | 19.6 | 2601 | 4 | US-09-627-650B-9 | Sequence 9, Appli  |
| 32 | 9 | 19.6 | 2601 | 4 | US-09-436-063C-9 | Sequence 9, Appli  |
| 33 | 8 | 17.4 | 18   | 1 | US-08-240-712-33 | Sequence 33, Appli |
| 34 | 8 | 17.4 | 18   | 1 | US-08-443-890-33 | Sequence 33, Appli |
| 35 | 8 | 17.4 | 54   | 1 | US-08-605-053-1  | Sequence 1, Appli  |
| 36 | 8 | 17.4 | 67   | 2 | US-09-010-928B-5 | Sequence 5, Appli  |
| 37 | 8 | 17.4 | 155  | 1 | US-08-209-747-15 | Sequence 15, Appli |
| 38 | 8 | 17.4 | 155  | 1 | US-08-458-298-15 | Sequence 15, Appli |
| 39 | 8 | 17.4 | 341  | 2 | US-08-209-521-11 | Sequence 11, Appli |
| 40 | 8 | 17.4 | 394  | 4 | US-09-144-914-4  | Sequence 4, Appli  |
| 41 | 8 | 17.4 | 400  | 1 | US-07-989-991A-2 | Sequence 2, Appli  |
| 42 | 8 | 17.4 | 400  | 3 | US-08-621-255-2  | Sequence 2, Appli  |
| 43 | 8 | 17.4 | 400  | 4 | US-09-352-574-2  | Sequence 2, Appli  |
| 44 | 8 | 17.4 | 407  | 2 | US-08-765-875-2  | Sequence 2, Appli  |
| 45 | 8 | 17.4 | 407  | 2 | US-08-765-875-6  | Sequence 6, Appli  |

ALIGNMENTS

RESULT 1  
US-09-009-816-4  
; Sequence 4, Application US/09009816  
; Patent No. 6436667  
; GENERAL INFORMATION:  
; APPLICANT: Gorman, Michael  
; APPLICANT: Permutt, M. Alan  
; APPLICANT: Inoue, Hiroshi  
; TITLE OF INVENTION: Human Nkx-6.1 Polypeptide-Encoding  
; TITLE OF INVENTION: Nucleotide Sequences  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bozicevic & Reed, LLP  
; STREET: 285 Hamilton Ave, Suite 200  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94301

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA: US/09/009,816  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Francis, Carol L  
; REGISTRATION NUMBER: 36,513  
; REFERENCE/DOCKET NUMBER: 9076/082CIP2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-327-3400  
; TELEFAX: 650-327-3231  
; TELETYPE:  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 223 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; US-09-009-816-4

Query Match 21.7%; Score 10; DB 4; Length 223;  
Best Local Similarity 13.3%; Pred. No. 2.2e+02;

| Matches  | 2;  | Conservative    | 0;  | Mismatches | 13; | Indels | 0; | Gaps | 0; |
|--|-----|-----------------|-----|------------|-----|--------|----|------|----|
| QY   | 4   | GXXXXXXX        | 18  |            |     |        |    |      |    |
| DB   | 122 | GSSSSSSSSASASSA | 136 |            |     |        |    |      |    |
| <p>RESULT 2</p> <p>US-09-009-816-2</p> <p>Sequence 2, Application US/09009816</p> <p>Patent No. 6436667</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: German, Michael</p> <p>APPLICANT: Permutt, M. Alan</p> <p>APPLICANT: Inoue, Hiroshi</p> <p>TITLE OF INVENTION: Human Nkx-6.1 Polypeptide-Encoding Nucleotide Sequences</p> <p>TITLE OF INVENTION: Nucleotide Sequences</p> <p>NUMBER OF SEQUENCES: 21</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: Bozicevic &amp; Reed, LLP</p> <p>STREET: 285 Hamilton Ave, Suite 200</p> <p>CITY: Palo Alto</p> <p>STATE: CA</p> <p>COUNTRY: USA</p> <p>ZIP: 94301</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Diskette</p> <p>COMPUTER: IBM Compatible</p> <p>OPERATING SYSTEM: DOS</p> <p>SOFTWARE: FastSeq for Windows Version 2.0</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/09/009,816</p> <p>FILING DATE:</p> <p>CLASSIFICATION:</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER:</p> <p>FILING DATE:</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: Francis, Carol L</p> <p>REGISTRATION NUMBER: 36,513</p> <p>REFERENCE/DOCKET NUMBER: 9076/082CIP2</p> <p>TELECOMMUNICATION INFORMATION:</p> <p>TELEPHONE: 650-327-3400</p> <p>TELEFAX: 650-327-3231</p> <p>TELEX:</p> <p>INFORMATION FOR SEQ ID NO: 2:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 367 amino acids</p> <p>TYPE: amino acid</p> <p>STRANDEDNESS: single</p> <p>TOPOLOGY: linear</p> <p>MOLECULE TYPE: protein</p> <p>FRAGMENT TYPE: internal</p> <p>US-09-009-816-2</p> |     |                 |     |            |     |        |    |      |    |
| <p>Query Match 21.7%; Score 10; DB 4; Length 367;</p> <p>Best Local Similarity 13.3%; Pred. No. 2.5e+02;</p> <p>Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;</p>  |     |                 |     |            |     |        |    |      |    |
| QY   | 4   | GXXXXXXX        | 18  |            |     |        |    |      |    |
| DB   | 122 | GSSSSSSSSASASSA | 136 |            |     |        |    |      |    |
| <p>RESULT 3</p> <p>US-08-864-038A-3</p> <p>Sequence 3, Application US/08864038A</p> <p>Patent No. 6001592</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Kunio NAKASHIMA et al.</p> <p>TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING</p> <p>TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING</p> <p>NUMBER OF SEQUENCES: 3</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: Incyte Pharmaceuticals, Inc.</p> <p>STREET: 3174 Porter Drive</p> <p>CITY: Palo Alto</p> <p>STATE: CA</p> <p>COUNTRY: USA</p> <p>ZIP: 94304</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Diskette</p> <p>COMPUTER: IBM Compatible</p>   |     |                 |     |            |     |        |    |      |    |
| <p>Query Match 21.7%; Score 10; DB 3; Length 738;</p> <p>Best Local Similarity 13.3%; Pred. No. 3.1e+02;</p> <p>Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;</p>  |     |                 |     |            |     |        |    |      |    |
| QY   | 4   | GXXXXXXX        | 18  |            |     |        |    |      |    |
| DB   | 351 | GAAAAAAAASASA   | 365 |            |     |        |    |      |    |
| <p>RESULT 4</p> <p>US-08-977-767-3</p> <p>Sequence 3, Application US/08977767</p> <p>Patent No. 5972684</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Bandman, Olga</p> <p>APPLICANT: Yue Henry</p> <p>APPLICANT: Greenwald, Sara</p> <p>APPLICANT: Corley, Neil C</p> <p>TITLE OF INVENTION: CARBONIC ANHYDRASE VIII</p> <p>NUMBER OF SEQUENCES: 3</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: Incyte Pharmaceuticals, Inc.</p> <p>STREET: 3174 Porter Drive</p> <p>CITY: Palo Alto</p> <p>STATE: CA</p> <p>COUNTRY: USA</p> <p>ZIP: 94304</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Diskette</p> <p>COMPUTER: IBM Compatible</p>   |     |                 |     |            |     |        |    |      |    |

```
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/977,767
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PP-0423 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1345 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1532042
US-08-977-767-3

Query Match 21.7%; Score 10; DB 2; Length 1345;
Best Local Similarity 13.3%; Pred. No. 3.8e+02;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXXA 18
|
Db 1322 GTAAAAAAXXXXXA 1336

RESULT 5
US-09-540-245A-16
; Sequence 16, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,245A
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 1381
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-540-245A-16

Query Match 21.7%; Score 10; DB 4; Length 1381;
Best Local Similarity 13.3%; Pred. No. 3.8e+02;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXXA 18
|
Db 684 GSTSTSASASASA 698

RESULT 6
US-09-627-650B-1

; Sequence 1, Application US/09627650B
; Patent No. 6406872
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and Methods Related Thereto
; FILE REFERENCE: 21101.000903
; CURRENT APPLICATION NUMBER: US/09/627,650B
; CURRENT FILING DATE: 2000-07-28
; PRIOR FILING DATE: 1999-11-08
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107,727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1652
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-1

Query Match 21.7%; Score 10; DB 4; Length 1652;
Best Local Similarity 13.3%; Pred. No. 4e+02;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXXA 18
|
Db 1612 GAAAAAAXXXXXA 1626

RESULT 7
US-09-436-063C-1
; Sequence 1, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and Methods Related Thereto
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436,063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1652
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-1

Query Match 21.7%; Score 10; DB 4; Length 1652;
Best Local Similarity 13.3%; Pred. No. 4e+02;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXXA 18
|
Db 1612 GAAAAAAXXXXXA 1626

RESULT 8
US-08-240-712-29
; Sequence 29, Application US/08240712
; Patent No. 5599907
; GENERAL INFORMATION:
; APPLICANT: ANDERSON, DAVID C.
; APPLICANT: MATHEWS, ANTONY JAMES
; APPLICANT: STETLER, GARY L.
; TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC HEMOGLOBINS
; NUMBER OF SEQUENCES: 35
```

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Browdy and Neimark  
;; STREET: 419 Seventh Street, N.W., Suite 300  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20004  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC Compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/240.712  
;; FILING DATE: 09-MAY-1994  
;; CLASSIFICATION: 530  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US92/09752  
;; FILING DATE: 13-MAY-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: COOPER, IVER P  
;; REGISTRATION NUMBER: 28,005  
;; REFERENCE/DOCKET NUMBER: ANDERSON=6  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-628-5197  
;; TELEFAX: 202-737-3528  
;; TELEX: 248633  
;;  
;; INFORMATION FOR SEQ ID NO: 29:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 182 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FEATURE:  
;; OTHER INFORMATION: /note= one or both of Gly  
;; OTHER INFORMATION: residues 131 and 132 can be absent; one or both  
;; OTHER INFORMATION: of Gly residues 147 and 148 can be absent  
;;  
US-08-240-712-29

Query Match 19.6%; Score 9; DB 1; Length 182;  
Best Local Similarity 13.3%; Pred. No. 9.3e+02;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXX 18  
|  
Db 131 GGAAAAA 145

RESULT 9  
US-08-443-890-29  
; Sequence 29, Application US/08443890  
; Patent No. 5739011  
; GENERAL INFORMATION:  
; APPLICANT: ANDERSON, DAVID C.  
; APPLICANT: MATHEWS, ANTONY JAMES  
; APPLICANT: STETLER, GARY L.  
; TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC  
; TITLE OF INVENTION: HEMOGLOBINS  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Browdy and Neimark  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/443,890  
;; FILING DATE: 31-MAY-1995  
;; CLASSIFICATION: 530  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/240,712  
;; FILING DATE: 09-MAY-1994  
;; APPLICATION NUMBER: PCT/US92/09752  
;; FILING DATE: 13-MAY-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: COOPER, IVER P  
;; REGISTRATION NUMBER: 28,005  
;; REFERENCE/DOCKET NUMBER: ANDERSON=6  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-628-5197  
;; TELEFAX: 202-737-3528  
;; TELEX: 248633  
;;  
;; INFORMATION FOR SEQ ID NO: 29:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 182 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FEATURE:  
;; OTHER INFORMATION: /note= one or both of Gly  
;; OTHER INFORMATION: residues 131 and 132 can be absent; one or both  
;; OTHER INFORMATION: of Gly residues 147 and 148 can be absent  
;;  
US-08-443-890-29

Query Match 19.6%; Score 9; DB 1; Length 182;  
Best Local Similarity 13.3%; Pred. No. 9.3e+02;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXX 18  
|  
Db 131 GGAAAAA 145

RESULT 10  
US-08-997-080-147  
; Sequence 147, Application US/08997080  
; Patent No. 5968524  
; GENERAL INFORMATION:  
; APPLICANT: WATSON, JAMES D.  
; APPLICANT: TAN, PAUL L.J.  
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-  
; NUMBER OF SEQUENCES: 194  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Law Offices of Ann W. Speckman  
; STREET: 2601 Elliott Avenue, Suite 4185  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/997,080  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sleath, Janet  
; REGISTRATION NUMBER: 37,007  
; REFERENCE/DOCKET NUMBER: 11000.1007  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-269-0565  
; TELEFAX: 206-269-0563





; TITLE OF INVENTION: Methods and Compounds for the Treatment  
; TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders  
; FILE REFERENCE: 11000.1007c1  
; CURRENT APPLICATION NUMBER: US/09/324,542  
; CURRENT FILING DATE: 1999-06-02 08/997,080  
; EARLIER APPLICATION NUMBER: US 08/997,080  
; EARLIER FILING DATE: 1997-12-23  
; NUMBER OF SEQ ID NOS: 194  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 147  
; LENGTH: 228  
; TYPE: PRT  
; ORGANISM: Mycobacterium vaccae  
US-09-324-542-147

Query Match 19.6%; Score 9; DB 4; Length 228;  
Best Local Similarity 13.3%; Pred. No. 1e+03;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

4 GXXXXXXXXXXXXXA 18  
|  
30 GGSTASSASTASSA 44

RESULT 14  
US-09-205-426-147  
; Sequence 147, Application US/09205426  
; Patent No. 6406704  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Tan, Paul L. J.  
; TITLE OF INVENTION: Compounds and Methods for Treatment and  
; TITLE OF INVENTION: Diagnosis of Mycobacterial Infections  
; FILE REFERENCE: 11000.1002c4  
; CURRENT APPLICATION NUMBER: US/09/205,426  
; CURRENT FILING DATE: 1998-12-04  
; EARLIER APPLICATION NUMBER: 09/095,855  
; EARLIER FILING DATE: 1998-06-11  
; EARLIER APPLICATION NUMBER: 08/997,362  
; EARLIER FILING DATE: 1997-12-23  
; EARLIER APPLICATION NUMBER: 08/873,970  
; EARLIER FILING DATE: 1997-06-12  
; EARLIER APPLICATION NUMBER: 08/705,347  
; EARLIER FILING DATE: 1996-08-29  
; NUMBER OF SEQ ID NOS: 208  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 147  
; LENGTH: 228  
; TYPE: PRT  
; ORGANISM: Mycobacterium vaccae  
US-09-205-426-147

Query Match 19.6%; Score 9; DB 4; Length 228;  
Best Local Similarity 13.3%; Pred. No. 1e+03;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GXXXXXXXXXXXXXA 18  
|  
Db 30 GGSTASSASTASSA 44

RESULT 15  
US-09-186-276B-34  
; Sequence 34, Application US/09186276B  
; Patent No. 6388173  
; GENERAL INFORMATION:  
; APPLICANT: Benfey, Philip  
; APPLICANT: DiLaurenzio, Laura  
; APPLICANT: Wysocka-Diller, Joanna  
; APPLICANT: Malamy, Jocelyn E.  
; APPLICANT: Pysch, Leonard  
; APPLICANT: Helariutta, Yrjo  
; TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses Thereof

; FILE REFERENCE: 5914-075-999  
; CURRENT APPLICATION NUMBER: US/09/186,276B  
; CURRENT FILING DATE: 1998-11-05  
; PRIOR APPLICATION NUMBER: 08/842,445  
; PRIOR FILING DATE: 1997-04-24  
; PRIOR APPLICATION NUMBER: 08/638,617  
; PRIOR FILING DATE: 1996-04-26  
; NUMBER OF SEQ ID NOS: 79  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 34  
; LENGTH: 277  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(277)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-186-276B-34

Query Match 19.6%; Score 9; DB 4; Length 277;  
Best Local Similarity 13.3%; Pred. No. 1.1e+03;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GXXXXXXXXXXXXXA 18  
|  
Db 28 GTTVTTTTAAAA 42

Search completed: January 8, 2003, 10:59:31  
Job time : 12.8103 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: January 8, 2003, 10:55:49 ; Search time 6.87931 seconds  
(without alignments)  
107.168 Million cell updates/sec

Title: US-09-003-869-4

Perfect score: 46

Sequence: 1 XX 38

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 118974 seqs, 19401057 residues

Total number of hits satisfying chosen parameters: 118974

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications-AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/FCI\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Query Match | Score | Length | ID | Description        |
|------------|-------------|-------|--------|----|--------------------|
| 1          | 10          | 21.7  | 220    | 10 | US-09-923-304-2    |
| 2          | 10          | 21.7  | 265    | 10 | US-09-925-302-559  |
| 3          | 10          | 21.7  | 330    | 9  | US-09-738-626-3767 |
| 4          | 10          | 21.7  | 914    | 9  | US-09-975-143-47   |
| 5          | 10          | 21.7  | 1076   | 9  | US-10-028-072-219  |
| 6          | 10          | 21.7  | 1463   | 9  | US-09-971-536-69   |
| 7          | 10          | 21.7  | 4440   | 12 | US-10-052-586-525  |
| 8          | 9           | 19.6  | 228    | 9  | US-10-051-643-147  |
| 9          | 9           | 19.6  | 329    | 10 | US-09-925-300-1406 |
| 10         | 9           | 19.6  | 480    | 10 | US-09-823-936-16   |
| 11         | 9           | 19.6  | 629    | 10 | US-09-833-760-429  |
| 12         | 9           | 19.6  | 723    | 12 | US-10-044-205A-32  |
| 13         | 9           | 19.6  | 739    | 12 | US-09-874-162A-5   |
| 14         | 9           | 19.6  | 785    | 12 | US-10-044-205A-31  |
| 15         | 9           | 19.6  | 1212   | 10 | US-09-981-353-16   |
| 16         | 9           | 19.6  | 1422   | 10 | US-09-735-933-1    |
| 17         | 8           | 17.4  | 32     | 10 | US-09-864-761-3920 |
| 18         | 8           | 17.4  | 120    | 10 | US-09-864-761-3793 |
| 19         | 8           | 17.4  | 130    | 10 | US-09-780-224A-9   |

|    |   |      |     |    |                    |                   |
|----|---|------|-----|----|--------------------|-------------------|
| 20 | 8 | 17.4 | 132 | 10 | US-09-780-224A-6   | Sequence 6, Appl  |
| 21 | 8 | 17.4 | 133 | 10 | US-09-780-224A-7   | Sequence 7, Appl  |
| 22 | 8 | 17.4 | 236 | 10 | US-09-925-297-649  | Sequence 649, App |
| 23 | 8 | 17.4 | 248 | 10 | US-09-925-301-114  | Sequence 114, Ap  |
| 24 | 8 | 17.4 | 352 | 9  | US-10-017-327-2    | Sequence 2, Appl  |
| 25 | 8 | 17.4 | 365 | 9  | US-09-738-626-5854 | Sequence 5854, Ap |
| 26 | 8 | 17.4 | 394 | 10 | US-09-939-484-4    | Sequence 4, Appl  |
| 27 | 8 | 17.4 | 394 | 10 | US-09-939-483-4    | Sequence 4, Appl  |
| 28 | 8 | 17.4 | 394 | 10 | US-09-798-584-15   | Sequence 15, Appl |
| 29 | 8 | 17.4 | 405 | 10 | US-09-799-777-38   | Sequence 38, Appl |
| 30 | 8 | 17.4 | 407 | 9  | US-09-841-730-25   | Sequence 25, Appl |
| 31 | 8 | 17.4 | 407 | 10 | US-09-454-540-2    | Sequence 2, Appl  |
| 32 | 8 | 17.4 | 407 | 10 | US-09-454-540-6    | Sequence 6, Appl  |
| 33 | 8 | 17.4 | 407 | 10 | US-09-859-894A-2   | Sequence 2, Appl  |
| 34 | 8 | 17.4 | 408 | 9  | US-09-813-398-33   | Sequence 33, Appl |
| 35 | 8 | 17.4 | 426 | 9  | US-09-738-626-3531 | Sequence 3531, Ap |
| 36 | 8 | 17.4 | 438 | 9  | US-09-738-626-6703 | Sequence 6703, Ap |
| 37 | 8 | 17.4 | 442 | 10 | US-09-749-728B-11  | Sequence 11, Appl |
| 38 | 8 | 17.4 | 446 | 10 | US-09-853-386-69   | Sequence 69, Appl |
| 39 | 8 | 17.4 | 580 | 9  | US-09-922-364A-2   | Sequence 2, Appl  |
| 40 | 8 | 17.4 | 580 | 9  | US-09-254-590-2    | Sequence 2, Appl  |
| 41 | 8 | 17.4 | 580 | 9  | US-10-115-695-2    | Sequence 24, Appl |
| 42 | 8 | 17.4 | 585 | 10 | US-09-817-774-24   | Sequence 70, Appl |
| 43 | 8 | 17.4 | 643 | 10 | US-09-853-386-70   | Sequence 57, Appl |
| 44 | 8 | 17.4 | 655 | 10 | US-09-205-658-57   | Sequence 57, Appl |
| 45 | 8 | 17.4 | 655 | 10 | US-09-844-353A-57  | Sequence 57, Appl |

## ALIGNMENTS

RESULT 1  
US-09-923-304-2  
; Sequence 2, Application US/09923304  
; Patent No. US20020081612A1  
; GENERAL INFORMATION:  
; APPLICANT: KATZ, RUTH  
; APPLICANT: JIANG, FENG  
; TITLE OF INVENTION: DETECTION AND DIAGNOSIS OF SMOKING RELATED CANCERS  
; FILE REFERENCE: UTSC:658US  
; CURRENT APPLICATION NUMBER: US/09/923,304  
; CURRENT FILING DATE: 2001-08-06  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-923-304-2

Query Match 21.7%; Score 10; DB 10; Length 220;  
Best Local Similarity 13.3%; Pred. No. 7.3;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXXA 18  
|  
DB 148 GTAAAAAAXXXXXXXA 162

RESULT 2  
US-09-925-302-559  
; Sequence 559, Application US/09925302  
; Patent No. US20020044941A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: P0104  
; CURRENT APPLICATION NUMBER: US/09/925,302  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05918  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 896  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 559  
LENGTH: 265  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (4)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-302-559

Query Match 21.7%; Score 10; DB 10; Length 265;  
Best Local Similarity 13.3%; Pred. No. 7.8; Indels 0; Gaps 0;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GXXXXXXXA 18  
193 GTAAAAA 207

RESULT 3  
US-09-738-626-3767  
Sequence 3767, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OKAZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 3767  
LENGTH: 330  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-3767

Query Match 21.7%; Score 10; DB 9; Length 330;  
Best Local Similarity 13.3%; Pred. No. 8.5; Indels 13; Indels 0; Gaps 0;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GXXXXXXXA 18  
58 GSSAASSTSSASSA 72

RESULT 4  
US-09-975-143-47  
Sequence 47, Application US/09975143  
Patent No. US20020155513A1  
GENERAL INFORMATION:  
APPLICANT: HSU, Daniel, K.  
APPLICANT: LIU, Fu-Tong  
APPLICANT: DOWLING, Christopher, A.  
TITLE OF INVENTION: CIRRHOTIC LIVER AND HEPATOCELLULAR CARCINOMA

FILE REFERENCE: DANHSU.001C1  
CURRENT APPLICATION NUMBER: US/09/975,143  
CURRENT FILING DATE: 2001-10-10  
PRIOR APPLICATION NUMBER: PCT/US00/08561  
PRIOR FILING DATE: 2000-03-29  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 47  
LENGTH: 914  
TYPE: PRT  
ORGANISM: human  
US-09-975-143-47

Query Match 21.7%; Score 10; DB 9; Length 914;  
Best Local Similarity 13.3%; Pred. No. 12; Indels 13; Indels 0; Gaps 0;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GXXXXXXXA 18  
785 GATTAATAAAAAA 799

RESULT 5  
US-10-028-072-219  
Sequence 219, Application US/10028072  
Publication No. US20030004311A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeForge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang  
TITLE OF INVENTION:  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/10/028,072  
CURRENT FILING DATE: 2001-12-19  
PRIOR APPLICATION NUMBER: 60/049911  
PRIOR FILING DATE: 1997-06-18  
PRIOR APPLICATION NUMBER: 60/056974  
PRIOR FILING DATE: 1997-08-26  
PRIOR APPLICATION NUMBER: 60/059113  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059115  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059117  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059122  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059184  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059263  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/059352  
PRIOR FILING DATE: 1997-09-19  
PRIOR APPLICATION NUMBER: 60/059588  
PRIOR FILING DATE: 1997-09-19  
PRIOR APPLICATION NUMBER: 60/059836  
PRIOR FILING DATE: 1997-09-24  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/062285

1 PRIOR FILING DATE: 1997-10-17  
2 PRIOR APPLICATION NUMBER: 60/062287  
3 PRIOR FILING DATE: 1997-10-17  
4 PRIOR APPLICATION NUMBER: 60/062814  
5 PRIOR FILING DATE: 1997-10-24  
6 PRIOR APPLICATION NUMBER: 60/062816  
7 PRIOR FILING DATE: 1997-10-24  
8 PRIOR APPLICATION NUMBER: 60/063045  
9 PRIOR FILING DATE: 1997-10-24  
10 PRIOR APPLICATION NUMBER: 60/063082  
11 PRIOR FILING DATE: 1997-10-31  
12 PRIOR APPLICATION NUMBER: 60/063127  
13 PRIOR FILING DATE: 1997-10-24  
14 PRIOR APPLICATION NUMBER: 60/063327  
15 PRIOR FILING DATE: 1997-10-27  
16 PRIOR APPLICATION NUMBER: 60/063329  
17 PRIOR FILING DATE: 1997-10-27  
18 PRIOR APPLICATION NUMBER: 60/063550  
19 PRIOR FILING DATE: 1997-10-28  
20 PRIOR APPLICATION NUMBER: 60/063561  
21 PRIOR FILING DATE: 1997-10-28  
22 PRIOR APPLICATION NUMBER: 60/063704  
23 PRIOR FILING DATE: 1997-10-29  
24 PRIOR APPLICATION NUMBER: 60/063733  
25 PRIOR FILING DATE: 1997-10-29  
26 PRIOR APPLICATION NUMBER: 60/063735  
27 PRIOR FILING DATE: 1997-10-29  
28 PRIOR APPLICATION NUMBER: 60/063738  
29 PRIOR FILING DATE: 1997-10-29  
30 PRIOR APPLICATION NUMBER: 60/063755  
31 PRIOR FILING DATE: 1997-10-17  
32 PRIOR APPLICATION NUMBER: 60/064248  
33 PRIOR FILING DATE: 1997-11-03  
34 PRIOR APPLICATION NUMBER: 60/064809  
35 PRIOR FILING DATE: 1997-11-07  
36 PRIOR APPLICATION NUMBER: 60/065186  
37 PRIOR FILING DATE: 1997-11-12  
38 PRIOR APPLICATION NUMBER: 60/065846  
39 PRIOR FILING DATE: 1997-11-17  
40 PRIOR APPLICATION NUMBER: 60/066364  
41 PRIOR FILING DATE: 1997-11-21  
42 PRIOR APPLICATION NUMBER: 60/066453  
43 PRIOR FILING DATE: 1997-11-24  
44 PRIOR APPLICATION NUMBER: 60/066511  
45 PRIOR FILING DATE: 1997-11-24  
46 PRIOR APPLICATION NUMBER: 60/066770  
47 PRIOR FILING DATE: 1997-11-24  
48 PRIOR APPLICATION NUMBER: 60/069212  
49 PRIOR FILING DATE: 1997-12-11  
50 PRIOR APPLICATION NUMBER: 60/069278  
51 PRIOR FILING DATE: 1997-12-11  
52 PRIOR APPLICATION NUMBER: 60/069334  
53 PRIOR FILING DATE: 1997-12-11  
54 PRIOR APPLICATION NUMBER: 60/069694  
55 PRIOR FILING DATE: 1997-12-16  
56 PRIOR APPLICATION NUMBER: 60/072320  
57 PRIOR FILING DATE: 1998-01-23  
58 PRIOR APPLICATION NUMBER: 60/073612  
59 PRIOR FILING DATE: 1998-02-04  
60 PRIOR APPLICATION NUMBER: 60/074086  
61 PRIOR FILING DATE: 1998-02-09  
62 PRIOR APPLICATION NUMBER: 60/074092  
63 PRIOR FILING DATE: 1998-02-09  
64 PRIOR APPLICATION NUMBER: 60/077791  
65 PRIOR FILING DATE: 1998-03-12  
66 PRIOR APPLICATION NUMBER: 60/078910  
67 PRIOR FILING DATE: 1998-03-20  
68 PRIOR APPLICATION NUMBER: 60/079294  
69 PRIOR FILING DATE: 1998-03-25  
70 PRIOR APPLICATION NUMBER: 60/079663  
71 PRIOR FILING DATE: 1998-02-27  
72 PRIOR APPLICATION NUMBER: 60/079728  
73 PRIOR FILING DATE: 1998-03-27  
74 PRIOR APPLICATION NUMBER: 60/080165  
75 PRIOR FILING DATE: 1998-03-31  
76 PRIOR APPLICATION NUMBER: 60/081203  
77 PRIOR FILING DATE: 1998-04-09  
78 PRIOR APPLICATION NUMBER: 60/081229  
79 PRIOR FILING DATE: 1998-04-09  
80 PRIOR APPLICATION NUMBER: 60/081695  
81 PRIOR FILING DATE: 1998-04-14  
82 PRIOR APPLICATION NUMBER: 60/081817  
83 PRIOR FILING DATE: 1998-04-15  
84 PRIOR APPLICATION NUMBER: 60/081818  
85 PRIOR FILING DATE: 1998-04-15  
86 PRIOR APPLICATION NUMBER: 60/082999  
87 PRIOR FILING DATE: 1998-04-24  
88 PRIOR APPLICATION NUMBER: 60/083322  
89 PRIOR FILING DATE: 1998-04-28  
90 PRIOR APPLICATION NUMBER: 60/083545  
91 PRIOR FILING DATE: 1998-04-29  
92 PRIOR APPLICATION NUMBER: 60/084600  
93 PRIOR FILING DATE: 1998-05-07  
94 PRIOR APPLICATION NUMBER: 60/084627  
95 PRIOR FILING DATE: 1998-05-07  
96 PRIOR APPLICATION NUMBER: 60/084637  
97 PRIOR FILING DATE: 1998-05-07  
98 PRIOR APPLICATION NUMBER: 60/085149  
99 PRIOR FILING DATE: 1998-05-12  
100 PRIOR APPLICATION NUMBER: 60/085323  
101 PRIOR FILING DATE: 1998-05-13  
102 PRIOR APPLICATION NUMBER: 60/085338  
103 PRIOR FILING DATE: 1998-05-13  
104 PRIOR APPLICATION NUMBER: 60/085339  
105 PRIOR FILING DATE: 1998-05-13  
106 PRIOR APPLICATION NUMBER: 60/085579  
107 PRIOR FILING DATE: 1998-05-15  
108 PRIOR APPLICATION NUMBER: 60/085697  
109 PRIOR FILING DATE: 1998-05-15  
110 PRIOR APPLICATION NUMBER: 60/085704  
111 PRIOR FILING DATE: 1998-05-15  
112 PRIOR APPLICATION NUMBER: 60/086414  
113 PRIOR FILING DATE: 1998-05-22  
114 PRIOR APPLICATION NUMBER: 60/086430  
115 PRIOR FILING DATE: 1998-05-22  
116 PRIOR APPLICATION NUMBER: 60/087106  
117 PRIOR FILING DATE: 1998-05-28  
118 PRIOR APPLICATION NUMBER: 60/088026  
119 PRIOR FILING DATE: 1998-06-04  
120 PRIOR APPLICATION NUMBER: 60/088730  
121 PRIOR FILING DATE: 1998-06-10  
122 PRIOR APPLICATION NUMBER: 60/088741  
123 PRIOR FILING DATE: 1998-06-10  
124 PRIOR APPLICATION NUMBER: 60/088810  
125 PRIOR FILING DATE: 1998-06-10  
126 PRIOR APPLICATION NUMBER: 60/088858  
127 PRIOR FILING DATE: 19/98-06-11  
128 PRIOR APPLICATION NUMBER: 60/089532  
129 PRIOR FILING DATE: 1998-06-17  
130 PRIOR APPLICATION NUMBER: 60/089599  
131 PRIOR FILING DATE: 1998-06-17  
132 PRIOR APPLICATION NUMBER: 60/089907  
133 PRIOR FILING DATE: 1998-06-18  
134 PRIOR APPLICATION NUMBER: 60/089947  
135 PRIOR FILING DATE: 1998-06-19  
136 PRIOR APPLICATION NUMBER: 60/090349  
137 PRIOR FILING DATE: 1998-06-23  
138 PRIOR APPLICATION NUMBER: 60/090429  
139 PRIOR FILING DATE: 1998-06-24  
140 PRIOR APPLICATION NUMBER: 60/090445  
141 PRIOR FILING DATE: 1998-06-24  
142 PRIOR APPLICATION NUMBER: 60/090538  
143 PRIOR FILING DATE: 1998-06-24  
144 PRIOR APPLICATION NUMBER: 60/090863  
145 PRIOR FILING DATE: 1998-06-26  
146 PRIOR APPLICATION NUMBER: 60/091360

; PRIOR FILING DATE: 1998-07-01  
; PRIOR APPLICATION NUMBER: 60/091519  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091982  
; PRIOR FILING DATE: 1998-07-07

Query Match 21.7%; Score 10; DB 9; Length 1076;  
Best Local Similarity 13.3%; Pred. No. 13;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18  
|  
Db 1016 GATTTAATTAATA 1030

RESULT 6  
US-09-971-536-69  
; Sequence 69, Application US/09971536  
; Patent No. US20020159976A1

GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Havukkala, Iikka  
; APPLICANT: Bloksberg, Leonard  
; APPLICANT: Lubbers, Mark  
; APPLICANT: Dekker, James  
; APPLICANT: Christensen, Anna  
; APPLICANT: Holland, Ross  
; APPLICANT: O'Toole, Paul  
; APPLICANT: Reid, Julian  
; APPLICANT: Coolbear, Timothy

; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods

; FILE REFERENCE: 1043c2  
; CURRENT APPLICATION NUMBER: US/09/971,536  
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238  
; PRIOR FILING DATE: 2000-08-08  
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160  
; PRIOR FILING DATE: 2001-08-08  
; NUMBER OF SEQ IDS: 83  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 69

; TYPE: PRT  
; ORGANISM: Lactobacillus rhamnosus  
; US-09-971-536-69

Query Match 21.7%; Score 10; DB 9; Length 1463;  
Best Local Similarity 13.3%; Pred. No. 14;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18  
|  
Db 1117 GAASSAASASRA 1131

RESULT 7  
US-10-052-586-525  
; Sequence 525, Application US/10052586  
; Patent No. US20020127584A1

GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.

; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C1  
; CURRENT APPLICATION NUMBER: US/10/052,586  
; CURRENT FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059266  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/063120  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063121  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063486  
; PRIOR FILING DATE: 1997-10-21  
; PRIOR APPLICATION NUMBER: 60/063540  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063541  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063544  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063564  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063734  
; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 60/063870  
; PRIOR FILING DATE: 1997-10-31  
; PRIOR APPLICATION NUMBER: 60/064103  
; PRIOR FILING DATE: 1997-10-31  
; PRIOR APPLICATION NUMBER: 60/065311  
; PRIOR FILING DATE: 1997-11-13  
; PRIOR APPLICATION NUMBER: 60/066120  
; PRIOR FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: 60/066466  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/066772  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/069335  
; PRIOR FILING DATE: 1997-12-11  
; PRIOR APPLICATION NUMBER: 60/069425  
; PRIOR FILING DATE: 1997-12-12  
; PRIOR APPLICATION NUMBER: 60/069870  
; PRIOR FILING DATE: 1997-12-17  
; PRIOR APPLICATION NUMBER: 60/068017  
; PRIOR FILING DATE: 1997-12-18  
; PRIOR APPLICATION NUMBER: 60/077450  
; PRIOR FILING DATE: 1998-03-10  
; PRIOR APPLICATION NUMBER: 60/077632  
; PRIOR FILING DATE: 1998-03-11  
; PRIOR APPLICATION NUMBER: 60/077649  
; PRIOR FILING DATE: 1998-03-11  
; PRIOR APPLICATION NUMBER: 60/078886  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/078939  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/079664  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: 60/079786  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: 60/080107  
; PRIOR FILING DATE: 1998-03-31  
; PRIOR APPLICATION NUMBER: 60/080194  
; PRIOR FILING DATE: 1998-03-31  
; PRIOR APPLICATION NUMBER: 60/080327  
; PRIOR FILING DATE: 1998-04-01  
; PRIOR APPLICATION NUMBER: 60/080333  
; PRIOR FILING DATE: 1998-04-01  
; PRIOR APPLICATION NUMBER: 60/081049  
; PRIOR FILING DATE: 1998-04-08  
; PRIOR APPLICATION NUMBER: 60/081070

|    |                                     |
|----|-------------------------------------|
| 1  | PRIOR FILING DATE: 1998-04-08       |
| 2  | PRIOR APPLICATION NUMBER: 60/081195 |
| 3  | PRIOR FILING DATE: 1998-04-09       |
| 4  | PRIOR APPLICATION NUMBER: 60/081038 |
| 5  | PRIOR FILING DATE: 1998-04-15       |
| 6  | PRIOR APPLICATION NUMBER: 60/082568 |
| 7  | PRIOR FILING DATE: 1998-04-21       |
| 8  | PRIOR APPLICATION NUMBER: 60/082569 |
| 9  | PRIOR FILING DATE: 1998-04-21       |
| 10 | PRIOR APPLICATION NUMBER: 60/082704 |
| 11 | PRIOR FILING DATE: 1998-04-22       |
| 12 | PRIOR APPLICATION NUMBER: 60/082797 |
| 13 | PRIOR FILING DATE: 1998-04-22       |
| 14 | PRIOR APPLICATION NUMBER: 60/083322 |
| 15 | PRIOR FILING DATE: 1998-04-28       |
| 16 | PRIOR APPLICATION NUMBER: 60/083495 |
| 17 | PRIOR FILING DATE: 1998-04-29       |
| 18 | PRIOR APPLICATION NUMBER: 60/083496 |
| 19 | PRIOR FILING DATE: 1998-04-29       |
| 20 | PRIOR APPLICATION NUMBER: 60/083499 |
| 21 | PRIOR FILING DATE: 1998-04-29       |
| 22 | PRIOR APPLICATION NUMBER: 60/083559 |
| 23 | PRIOR FILING DATE: 1998-04-29       |
| 24 | PRIOR APPLICATION NUMBER: 60/084366 |
| 25 | PRIOR FILING DATE: 1998-05-05       |
| 26 | PRIOR APPLICATION NUMBER: 60/084414 |
| 27 | PRIOR FILING DATE: 1998-05-06       |
| 28 | PRIOR APPLICATION NUMBER: 60/084639 |
| 29 | PRIOR FILING DATE: 1998-05-07       |
| 30 | PRIOR APPLICATION NUMBER: 60/084640 |
| 31 | PRIOR FILING DATE: 1998-05-07       |
| 32 | PRIOR APPLICATION NUMBER: 60/084643 |
| 33 | PRIOR FILING DATE: 1998-05-07       |
| 34 | PRIOR APPLICATION NUMBER: 60/085573 |
| 35 | PRIOR FILING DATE: 1998-05-15       |
| 36 | PRIOR APPLICATION NUMBER: 60/085579 |
| 37 | PRIOR FILING DATE: 1998-05-15       |
| 38 | PRIOR APPLICATION NUMBER: 60/085580 |
| 39 | PRIOR FILING DATE: 1998-05-15       |
| 40 | PRIOR APPLICATION NUMBER: 60/085582 |
| 41 | PRIOR FILING DATE: 1998-05-15       |
| 42 | PRIOR APPLICATION NUMBER: 60/085700 |
| 43 | PRIOR FILING DATE: 1998-05-15       |
| 44 | PRIOR APPLICATION NUMBER: 60/086023 |
| 45 | PRIOR FILING DATE: 1998-05-18       |
| 46 | PRIOR APPLICATION NUMBER: 60/086392 |
| 47 | PRIOR FILING DATE: 1998-05-22       |
| 48 | PRIOR APPLICATION NUMBER: 60/086486 |
| 49 | PRIOR FILING DATE: 1998-05-22       |
| 50 | PRIOR APPLICATION NUMBER: 60/087098 |
| 51 | PRIOR FILING DATE: 1998-05-28       |
| 52 | PRIOR APPLICATION NUMBER: 60/087208 |
| 53 | PRIOR FILING DATE: 1998-05-28       |
| 54 | PRIOR APPLICATION NUMBER: 60/087609 |
| 55 | PRIOR FILING DATE: 1998-06-02       |
| 56 | PRIOR APPLICATION NUMBER: 60/087759 |
| 57 | PRIOR FILING DATE: 1998-06-02       |
| 58 | PRIOR APPLICATION NUMBER: 60/087827 |
| 59 | PRIOR FILING DATE: 1998-06-03       |
| 60 | PRIOR APPLICATION NUMBER: 60/088025 |
| 61 | PRIOR FILING DATE: 1998-06-04       |
| 62 | PRIOR APPLICATION NUMBER: 60/088028 |
| 63 | PRIOR FILING DATE: 1998-06-04       |
| 64 | PRIOR APPLICATION NUMBER: 60/088029 |
| 65 | PRIOR FILING DATE: 1998-06-04       |
| 66 | PRIOR APPLICATION NUMBER: 60/088033 |
| 67 | PRIOR FILING DATE: 1998-06-04       |
| 68 | PRIOR APPLICATION NUMBER: 60/088167 |
| 69 | PRIOR FILING DATE: 1998-06-05       |
| 70 | PRIOR APPLICATION NUMBER: 60/088202 |
| 71 | PRIOR FILING DATE: 1998-06-05       |
| 72 | PRIOR APPLICATION NUMBER: 60/088212 |
| 73 | PRIOR FILING DATE: 1998-06-05       |

Query Match 21.7%: Score 10; DB 12; Length 4440;

```

Query Match      21.7%; Score 10; DB 12; Length 4440;
Best Local Similarity 13.3%; Pred. No. 21;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

0v 4 GXXXXXXXXXXXXXA 18

DB 4240 GTATATTATTA 4254

```

RESULT 8
US-10-051-643-147
> Sequence 147, Application US/10051643
> Publication NO. US20020197265A1
> GENERAL INFORMATION:
> APPLICANT: Watson, James D.
> APPLICANT: Tan, Paul L. J.
> TITLE OF INVENTION: Methods and Compounds
> TITLE OF INVENTION: of immunologically-M
> TITLE OF INVENTION: system using Mycobact
> TITLE OF INVENTION: 10000862
> CURRENT FILING DATE: 2002-01-18
> CURRENT FILING DATE: 2002-01-18
> PRIOR APPLICATION NUMBER: US09/156,181
> PRIOR FILING DATE: 1998-09-17
> PRIOR APPLICATION NUMBER: US 08/956,624
> PRIOR FILING DATE: 1997-12-23
> NUMBER OF SEQ ID NOS: 208
> SOFTWARE: FastSeq for Windows Version 3.0.
> SEQ ID NO 147
> LENGTH: 228
> TYPE: PRT

```





Db 77 GLSAAAAAAAAAAAA 91

RESULT 13  
US-09-874-162A-5  
; Sequence 5, Application US/09874162A  
; Patent No. US20020155452A1

; GENERAL INFORMATION:

; APPLICANT: Koontz, Jason  
; APPLICANT: Sklar, Jeffrey  
; TITLE OF INVENTION: FUSION OF JAZF1 AND JAZ2 GENES IN  
; FILE REFERENCE: 05311-024001  
; CURRENT APPLICATION NUMBER: US/09/874,162A

; CURRENT FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: US 60/209,093

; PRIOR FILING DATE: 2000-06-02  
; NUMBER OF SEQ ID NOS: 23

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5

; LENGTH: 739

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-874-162A-5

Query Match 19.6%; Score 9; DB 9; Length 739;  
Best Local Similarity 13.3%; Pred. No. 1e+02;

Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18

Db 50 GSYSSSSSSAAAAA 64

RESULT 14

US-10-044-205A-31

; Sequence 31, Application US/10044205A  
; Patent No. US20020123464A1

; GENERAL INFORMATION:

; APPLICANT: KAPILLER-LIBERMANN, Rosana  
; APPLICANT: BANDARU, Rajasekhar

; TITLE OF INVENTION: 69087, 15821, and 15418, Methods and Compositions of Human Prote

; FILE REFERENCE: 10147-52U1  
; CURRENT APPLICATION NUMBER: US/10/044,205A

; CURRENT FILING DATE: 2002-04-19  
; PRIOR APPLICATION NUMBER: US 60/242,428

; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: US 60/241,884

; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: US 60/241,877

; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 31

; LENGTH: 796

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-044-205A-31

Query Match 19.6%; Score 9; DB 12; Length 796;  
Best Local Similarity 13.3%; Pred. No. 1e+02;

Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18

Db 150 GLSAAAAAAAAAAAAA 164

RESULT 15

US-09-981-353-16

; Sequence 16, Application US/09981353  
; Patent No. US20020160382A1

; GENERAL INFORMATION:  
; APPLICANT: Lasek, Amy W.  
; APPLICANT: Jones, David A.  
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER  
; FILE REFERENCE: PA-0038 US  
; CURRENT APPLICATION NUMBER: US/09/981,353  
; CURRENT FILING DATE: 2001-10-11  
; NUMBER OF SEQ ID NOS: 194  
; SOFTWARE: PERL Program  
; SEQ ID NO 16  
; LENGTH: 1212  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc-feature  
; FEATURE:  
; OTHER INFORMATION: Incyte ID No. US20020160382A1 1285632CD1  
US-09-981-353-16

Query Match 19.6%; Score 9; DB 9; Length 1212;  
Best Local Similarity 13.3%; Pred. No. 1.2e+02;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18

Db 91 GRAAAAAAAAAAAAAA 105

Search completed: January 8, 2003, 11:01:14  
Job time: 7.87931 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2003, 10:58:35 ; Search time 210.966 Seconds  
(without alignments)  
116.132 Million cell updates/sec

Title: US-09-003-869-4  
Perfect score: 46  
Sequence: 1 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 38

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

11 number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

```
1: /cgn2_6/ptodata1/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata1/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata1/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata1/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata1/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata1/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata1/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata1/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata1/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata1/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata1/paa/US097_COMB.pep.*
22: /cgn2_6/ptodata1/paa/US098_COMB.pep.*
23: /cgn2_6/ptodata1/paa/US099_COMB.pep.*
24: /cgn2_6/ptodata1/paa/US100_COMB.pep.*
25: /cgn2_6/ptodata1/paa/US101_COMB.pep.*
26: /cgn2_6/ptodata1/paa/US102_COMB.pep.*
27: /cgn2_6/ptodata1/paa/US60_COMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Match | Length | DB ID                  | Description       |
|------------|-------|-------|--------|------------------------|-------------------|
| 1          | 10    | 21.7  | 96     | 1 PCT-US01-08631-47186 | Sequence 47186, A |
| 2          | 10    | 21.7  | 100    | 11 US-08-715-713-16    | Sequence 16, Appl |
| 3          | 10    | 21.7  | 142    | 1 PCT-US01-14827-11001 | Sequence 11001, A |
| 4          | 10    | 21.7  | 148    | 21 US-09-733-089-9201  | Sequence 9201, Ap |
| 5          | 10    | 21.7  | 148    | 22 US-09-816-660-9201  | Sequence 9201, Ap |
| 6          | 10    | 21.7  | 150    | 1 PCT-US01-32045-64    | Sequence 64, Appl |

```

162 16 US-09-270-767-42531
173 20 US-09-649-238-2
21 21.7 184 21 US-09-733-089-20463
21 21.7 187 22 US-09-816-660-20463
21 21.7 187 22 US-09-252-991A-17564
21 21.7 191 21 US-09-733-089-17611
21 21.7 191 22 US-09-816-660-17611
21 21.7 211 21 US-09-791-537-122363
21 21.7 213 18 US-09-489-039A-11130
21 21.7 220 1 PCT-US02-10824-153
21 21.7 220 23 US-09-923-304-2
21 21.7 222 20 US-09-602-777A-380
21 21.7 222 20 US-09-603-124B-450
21 21.7 222 20 US-09-603-124B-452
21 21.7 224 10 US-08-626-831-3
21 21.7 228 21 US-09-708-427-59912
21 21.7 265 1 PCT-US00-05918-559
21 21.7 265 23 US-09-925-302-559
21 21.7 267 1 PCT-US01-14826-369
21 21.7 288 24 US-10-029-386-33741
21 21.7 300 1 PCT-US01-14826-801
21 21.7 304 1 PCT-US02-21179-1
21 21.7 305 21 US-09-791-537-106611
21 21.7 314 21 US-09-791-537-5230
21 21.7 314 21 US-09-791-537-29243
21 21.7 315 17 US-09-373-662-16
21 21.7 327 20 US-09-614-150-27588
21 21.7 327 27 US-60-173-464-23012
21 21.7 327 27 US-60-191-637-27817
21 21.7 327 27 US-60-191-681-22426
21 21.7 330 21 US-09-738-626-3767
21 21.7 333 23 US-09-935-625-12556
21 21.7 336 19 US-09-513-996A-17522
21 21.7 336 23 US-09-995-938A-6
21 21.7 336 23 US-09-995-938A-7
21 21.7 348 27 US-60-161-932-1753
21 21.7 350 1 PCT-US02-29636-489
21 21.7 350 1 PCT-US02-29964-791
21 21.7 353 21 US-09-791-537-119027

```

#### ALIGNMENTS

```

RESULT 1
PCT-US01-08631-47186
; Sequence 47186, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 47186
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (2)..(17)
; OTHER INFORMATION: POLLEN ALLERGEN POA PI SIGNATURE domain identified by
; OTHER INFORMATION: EMATRIX, accession number PR00833H, p-value=8.375e-10, raw sco
; OTHER INFORMATION: 2.30
PCT-US01-08631-47186

```

Query Match 21.7%; Score 10; DB 1; Length 96;  
Best Local Similarity 13.3%; Pred. No. 6.3e+02;

Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18  
|  
Db 2 GTAAAAA 16

## RESULT 2

US-08-715-713-16  
; Sequence 16, Application US/08715713  
; GENERAL INFORMATION:  
; APPLICANT: Innis, Jeffrey W.  
; APPLICANT: Mortlock, Douglas P.  
; APPLICANT: Nelson, Matt  
; TITLE OF INVENTION: PROMOTER CAPTURE  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/715.713  
; FILING DATE: 19-SEP-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carroll, Peter G.  
; REGISTRATION NUMBER: 32,837  
; REFERENCE/DOCKET NUMBER: UM-02402  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-715-713-16

Query Match 21.7%; Score 10; DB 11; Length 100;  
Best Local Similarity 13.3%; Pred. No. 6.5e+02;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18  
|  
Db 37 GAAAAA 51

## RESULT 3

PCT-US01-14827-11001  
; Sequence 11001, Application PC/TUS0114827  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 21272-104  
; CURRENT APPLICATION NUMBER: PCT/US01/14827  
; CURRENT FILING DATE: 2001-05-16  
; PRIOR APPLICATION NUMBER: 09/577,408  
; PRIOR FILING DATE: 2000-05-18  
; NUMBER OF SEQ ID NOS: 16102  
; SOFTWARE: Custom  
; SEQ ID NO 11001  
; LENGTH: 142  
; TYPE: PRT

; ORGANISM: Homo sapiens  
PCT-US01-14827-11001

Query Match 21.7%; Score 10; DB 1; Length 142;  
Best Local Similarity 13.3%; Pred. No. 7.5e+02;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18  
|  
Db 4 GSTTASAAA 18

## RESULT 4

US-09-733-089-9201  
; Sequence 9201, Application US/09733089  
; GENERAL INFORMATION:  
; APPLICANT: Dotson, Stanton B.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Lutfiyya, Linda L.  
; APPLICANT: McIninch, James  
; APPLICANT: Wu, Wei  
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
; FILE REFERENCE: 38-21(15300)D  
; CURRENT APPLICATION NUMBER: US/09/733,089  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 09/474,435  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: US 09/654,617  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: US 09/620,392  
; PRIOR FILING DATE: 2000-07-19  
; NUMBER OF SEQ ID NOS: 24143  
; SEQ ID NO 9201  
; LENGTH: 148  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-09-733-089-9201

Query Match 21.7%; Score 10; DB 21; Length 148;  
Best Local Similarity 13.3%; Pred. No. 7.7e+02;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18  
|  
Db 59 GSSSSAAAAA 73

## RESULT 5

US-09-816-660-9201  
; Sequence 9201, Application US/09816660  
; GENERAL INFORMATION:  
; APPLICANT: Dotson, Stanton B.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Lutfiyya, Linda L.  
; APPLICANT: McIninch, James  
; APPLICANT: Wu, Wei  
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
; FILE REFERENCE: 38-21(15300)D  
; CURRENT APPLICATION NUMBER: US/09/816,660  
; CURRENT FILING DATE: 2001-03-26  
; PRIOR APPLICATION NUMBER: US 09/474,435  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: US 09/654,617  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: US 09/733,089  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: )  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: US 09/620,392

; PRIOR FILING DATE: 2000-07-19  
; NUMBER OF SEQ ID NOS: 24143  
; SEQ ID NO 9201  
; LENGTH: 148  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-09-816-660-9201

Query Match 21.7%; Score 10; DB 22; Length 148;  
Best Local Similarity 13.3%; Pred. No. 7.7e+02;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18  
|  
DB 59 GSSSSAAAAA 73

RESULT 6  
Sequence 64, Application PC/TUS0132045

GENERAL INFORMATION:  
; APPLICANT: Gish, Kurt C.  
; APPLICANT: Mack, David H.  
; APPLICANT: Wilson, Keith E.  
; APPLICANT: Afar, Daniel  
; APPLICANT: Hevezi, Peter  
; APPLICANT: Eos Biotechnology, Inc.  
; TITLE OF INVENTION: Methods of Diagnosis of Prostate Cancer, Compositions  
; TITLE OF INVENTION: and Methods of Screening for Modulators of Prostate  
; FILE REFERENCE: 018501-004200PC  
; CURRENT APPLICATION NUMBER: PCT/US01/32045  
; CURRENT FILING DATE: 2002-08-22  
; PRIOR APPLICATION NUMBER: US 09/687,576  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: US 09/733,288  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: US 09/733,742  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: US 60/263,957  
; PRIOR FILING DATE: 2001-01-24  
; PRIOR APPLICATION NUMBER: US 60/276,791  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: US 60/276,888  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: US 60/281,922  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: US 60/286,214  
; PRIOR FILING DATE: 2001-04-24  
; PRIOR APPLICATION NUMBER: US 09/847,046  
; PRIOR FILING DATE: 2001-04-30  
; PRIOR APPLICATION NUMBER: US 60/288,589  
; PRIOR FILING DATE: 2001-05-04  
; NUMBER OF SEQ ID NOS: 296  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 64  
; LENGTH: 150  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US01-32045-64

Query Match 21.7%; Score 10; DB 1; Length 150;  
Best Local Similarity 13.3%; Pred. No. 7.7e+02;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18  
|  
DB 126 GTAAAAA 140

RESULT 7  
US-09-270-767-42531  
; Sequence 42531, Application US/09270767

GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 42531  
; LENGTH: 162  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
; OTHER INFORMATION: Xaa means any amino acid  
US-09-270-767-42531

Query Match 21.7%; Score 10; DB 16; Length 162;  
Best Local Similarity 13.3%; Pred. No. 8e+02;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18  
|  
DB 99 GAAATSAATSA 113

RESULT 8  
US-09-649-238-2  
; Sequence 2, Application US/09649238  
; GENERAL INFORMATION:  
; APPLICANT: PFAFF, SAMUEL L.  
; TITLE OF INVENTION: MOTOR NEURON SPECIFIC PROMOTER AND USES THEREOF  
; FILE REFERENCE: SALK3010-1  
; CURRENT APPLICATION NUMBER: US/09/649,238  
; CURRENT FILING DATE: 2000-08-28  
; PRIOR APPLICATION NUMBER: 60/151,184  
; PRIOR FILING DATE: 1999-08-26  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 173  
; TYPE: PRT  
; ORGANISM: Murine sp.  
US-09-649-238-2

Query Match 21.7%; Score 10; DB 20; Length 173;  
Best Local Similarity 13.3%; Pred. No. 8.2e+02;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18  
|  
DB 120 GAAAAA 134

RESULT 9  
US-09-733-089-20463  
; Sequence 20463, Application US/09733089  
; GENERAL INFORMATION:  
; APPLICANT: Dotson, Stanton B.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Lufiyva, Linda L.  
; APPLICANT: McIninch, James  
; APPLICANT: Wu, Wei  
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
; FILE REFERENCE: 38-21(15300)D  
; CURRENT APPLICATION NUMBER: US/09/733,089  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 09/474,435  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: US 09/654,617  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: US 09/620,392

```
; PRIORITY FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 24143
; SEQ ID NO 20463
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(184)
; OTHER INFORMATION: unsure at all xaa locations
US-09-733-089-20463

Query Match      21.7%; Score 10; DB 21; Length 184;
Best Local Similarity 13.3%; Pred. No. 8.4e+02;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXXA 18
DB 72 GSSSSSTSTSSSSA 86

RESULT 10
US-09-816-660-20463
; Sequence 20463, Application US/09816660
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyva, Linda L.
; APPLICANT: McIninch, James
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-21(15300)D
; CURRENT APPLICATION NUMBER: US/09/816,660
; PRIORITY FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US/09/474,435
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: US/09/654,617
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US/09/733,089
; PRIOR FILING DATE: 2000-12-11
; PRIOR APPLICATION NUMBER: )
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US/09/620,392
; PRIOR FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 24143
; SEQ ID NO 20463
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(184)
; OTHER INFORMATION: unsure at all xaa locations
US-09-816-660-20463

Query Match      21.7%; Score 10; DB 22; Length 184;
Best Local Similarity 13.3%; Pred. No. 8.4e+02;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXXA 18
DB 72 GSSSSSTSTSSSSA 86

RESULT 11
US-09-252-991A-17564
; Sequence 17564, Application US/09252991A
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
```

```
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17564
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17564

Query Match      21.7%; Score 10; DB 16; Length 187;
Best Local Similarity 13.3%; Pred. No. 8.5e+02;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXXA 18
DB 63 GSAAAASTSTASTSA 77

RESULT 12
US-09-733-089-17611
; Sequence 17611, Application US/09733089
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyva, Linda L.
; APPLICANT: McIninch, James
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-21(15300)D
; CURRENT APPLICATION NUMBER: US/09/733,089
; CURRENT FILING DATE: 2000-12-11
; PRIOR APPLICATION NUMBER: US/09/474,435
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: US/09/654,617
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US/09/620,392
; PRIOR FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 24143
; SEQ ID NO 17611
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-733-089-17611

Query Match      21.7%; Score 10; DB 21; Length 191;
Best Local Similarity 13.3%; Pred. No. 8.6e+02;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXXA 18
DB 32 GSSSSSTSTSSSSA 46

RESULT 13
US-09-816-660-17611
; Sequence 17611, Application US/09816660
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyva, Linda L.
; APPLICANT: McIninch, James
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-21(15300)D
```

```
; CURRENT APPLICATION NUMBER: US/09/816,660
; CURRENT FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US 09/474,435
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 09/733,089
; PRIOR FILING DATE: 2000-12-11
; PRIOR APPLICATION NUMBER: ) US 09/684,016
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/620,392
; PRIOR FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 24143
; SEQ ID NO 17611
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-816-660-17611

Query Match      21.7%; Score 10; DB 22; Length 191;
Best Local Similarity 13.3%; Pred. No. 8.6e+02;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GXXXXXXXXXXXXXA 18
Db 32 GSSSSSTSTSSSA 46

RESULT 14
US-09-791-122363
; Sequence 122363, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 122363
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-791-537-122363

Query Match      21.7%; Score 10; DB 21; Length 211;
Best Local Similarity 13.3%; Pred. No. 9e+02;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GXXXXXXXXXXXXXA 18
Db 40 GTSASASSAAAAA 54

RESULT 15
US-09-489-039A-11130
; Sequence 11130, Application US/09489039A
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 11130
; LENGTH: 213
; TYPE: PRT
```

```
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11130

Query Match      21.7%; Score 10; DB 18; Length 213;
Best Local Similarity 13.3%; Pred. No. 9e+02;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GXXXXXXXXXXXXXA 18
Db 13 GTTTAASSAASAASA 27

Search completed: January 8, 2003, 11:12:07
Job time : 212.966 secs
```





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2003, 10:58:59 ; Search time 10.8103 Seconds  
(without alignments)  
249.151 Million cell updates/sec

Title: US-09-003-869-4  
Perfect score: 46  
Sequence: 1 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 38

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262194 seqs, 70879160 residues

1 number of hits satisfying chosen parameters: 262194

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents\_AA\_New\*

- 1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/paa/US05\_NEW\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query | Length | DB ID                | Description        |
|------------|-------|-------|--------|----------------------|--------------------|
| 1          | 10    | 21.7  | 218    | US-09-513-999C-4836  | Sequence 4836, Ap  |
| 2          | 10    | 21.7  | 218    | US-09-513-999C-4840  | Sequence 4840, Ap  |
| 3          | 10    | 21.7  | 225    | US-09-513-999C-4839  | Sequence 4839, Ap  |
| 4          | 10    | 21.7  | 252    | PCT-US02-18256-17    | Sequence 17, Appl  |
| 5          | 10    | 21.7  | 299    | PCT-US02-29560-320   | Sequence 320, Appl |
| 6          | 10    | 21.7  | 373    | US-09-724-676-78469  | Sequence 78469, A  |
| 7          | 10    | 21.7  | 373    | US-09-724-676-78470  | Sequence 78470, A  |
| 8          | 10    | 21.7  | 373    | US-09-724-676A-78469 | Sequence 78469, A  |
| 9          | 10    | 21.7  | 373    | US-09-724-676A-78470 | Sequence 78470, A  |
| 10         | 10    | 21.7  | 399    | US-09-724-676-78468  | Sequence 78468, A  |
| 11         | 10    | 21.7  | 399    | US-09-724-676A-78468 | Sequence 78468, A  |
| 12         | 10    | 21.7  | 462    | US-10-270-333-33     | Sequence 33, Appl  |
| 13         | 10    | 21.7  | 1076   | US-10-131-813A-219   | Sequence 219, App  |
| 14         | 10    | 21.7  | 1076   | US-10-131-819A-219   | Sequence 219, App  |
| 15         | 10    | 21.7  | 1076   | US-10-131-823A-219   | Sequence 219, App  |
| 16         | 10    | 21.7  | 1076   | US-10-131-824A-219   | Sequence 219, App  |
| 17         | 10    | 21.7  | 1076   | US-10-131-826A-219   | Sequence 219, App  |
| 18         | 10    | 21.7  | 1076   | US-10-131-829A-219   | Sequence 219, App  |
| 19         | 10    | 21.7  | 1076   | US-10-125-926A-219   | Sequence 219, App  |
| 20         | 10    | 21.7  | 1076   | US-10-127-829A-219   | Sequence 219, App  |
| 21         | 10    | 21.7  | 1076   | US-10-127-831A-219   | Sequence 219, App  |
| 22         | 10    | 21.7  | 1076   | US-10-127-835A-219   | Sequence 219, App  |
| 23         | 10    | 21.7  | 1076   | US-10-127-837A-219   | Sequence 219, App  |
| 24         | 10    | 21.7  | 1076   | US-10-127-842A-219   | Sequence 219, App  |
| 25         | 10    | 21.7  | 1076   | US-10-127-850A-219   | Sequence 219, App  |
| 26         | 10    | 21.7  | 1076   | US-10-127-901A-219   | Sequence 219, App  |

ALIGNMENTS

RESULT 1

US-09-513-999C-4836  
; Sequence 4836, Application US/09513999C  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; FILE REFERENCE: 59.US2.REC  
; CURRENT APPLICATION NUMBER: US/09/513.999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 4836  
; LENGTH: 218  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: 197  
; OTHER INFORMATION: Xaa=Leu or Pro or Gln or Arg  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: 218  
; OTHER INFORMATION: Xaa=Lys or Asn  
US-09-513-999C-4836

Query Match 21.7%; Score 10; DB 5; Length 218;

Best Local Similarity 13.3%; Pred. No. 4e-155;

Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXX 18

Db 148 GTAAAAA 162

RESULT 2

US-09-513-999C-4840  
; Sequence 4840, Application US/09513999C  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; FILE REFERENCE: 59.US2.REC  
; CURRENT APPLICATION NUMBER: US/09/513.999C  
; CURRENT FILING DATE: 2000-02-24

```
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 3681
; SOFTWARE: Patent.pm
; SEQ ID NO 4840
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 69
; OTHER INFORMATION: Xaa=His or Gln
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 105
; OTHER INFORMATION: Xaa=His or Ile or Lys or Leu or Met or Asn or Pro or Gln or Arg
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 197
; OTHER INFORMATION: Xaa=Leu or Pro or Gln or Arg
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 218
; OTHER INFORMATION: Xaa=Lys or Asn
; US-09-513-999C-4840

Query Match      21.7%; Score 10; DB 5; Length 218;
Best Local Similarity 13.3%; Pred. No. 4e-155;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY      4 GXXXXXXXXXXXXXA 18
Db      148 GTAAAAAAXXXXXX 162

RESULT 3
US-09-513-999C-4839
; Sequence 4839, Application US/09513999C
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59-US2.REG
; CURRENT APPLICATION NUMBER: US/09/513.999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 3681
; SOFTWARE: Patent.pm
; SEQ ID NO 4839
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 204
; OTHER INFORMATION: Xaa=Leu or Pro or Gln or Arg
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 225
; OTHER INFORMATION: Xaa=Lys or Asn
; US-09-513-999C-4839

Query Match      21.7%; Score 10; DB 5; Length 225;
Best Local Similarity 13.3%; Pred. No. 4e-155;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY      4 GXXXXXXXXXXXXXA 18
Db      148 GTAAAAAAXXXXXX 162
```

```
RESULT 4
PCT-US02-18256-17
; Sequence 17, Application PC/TUS0218256
; GENERAL INFORMATION:
; APPLICANT: Don A. Roth
; APPLICANT: Randolph V. Lewis
; APPLICANT: The University of Wyoming
; TITLE OF INVENTION: Expression of Spider Silk Proteins in Higher Plants
; FILE REFERENCE: UWVO 02-004
; CURRENT APPLICATION NUMBER: PCT/US02/18256
; CURRENT FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: 60/296,184
; PRIOR FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Zorocrates sp.
; PCT-US02-18256-17

Query Match      21.7%; Score 10; DB 1; Length 252;
Best Local Similarity 13.3%; Pred. No. 4e-155;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY      4 GXXXXXXXXXXXXXA 18
Db      25 GAAAAAAXXXXXX 39

RESULT 5
PCT-US02-29560-320
; Sequence 320, Application PC/TUS0229560
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Gish, Kurt C.
; APPLICANT: Hevezl, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Wilson, Keith E.
; APPLICANT: Zlotnik, Albert
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-002710PC
; CURRENT APPLICATION NUMBER: PCT/US02/29560
; CURRENT FILING DATE: 2025-11-01
; PRIOR APPLICATION NUMBER: US 60/323,469
; PRIOR FILING DATE: 2001-09-17
; NUMBER OF SEQ ID NOS: 412
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 320
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
; PCT-US02-29560-320

Query Match      21.7%; Score 10; DB 1; Length 299;
Best Local Similarity 13.3%; Pred. No. 4.1e-155;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY      4 GXXXXXXXXXXXXXA 18
Db      276 GTAAAAAAXXXXXX 290

RESULT 6
US-09-724-676-78469
; Sequence 78469, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
```

```
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 78469
; LENGTH: 373
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-78469

Query Match      21.7%; Score 10; DB 5; Length 373;
Best Local Similarity 13.3%; Pred. No. 4.1e-155;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXX 18
|
Db 45 GAAASSASAAAAASAA 59

RESULT 7
US-09-724-676-78470
; Sequence 78470, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 78470
; LENGTH: 373
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-78470

Query Match      21.7%; Score 10; DB 5; Length 373;
Best Local Similarity 13.3%; Pred. No. 4.1e-155;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXX 18
|
Db 45 GAAASSASAAAAASAA 59

US-09-724-676-78468
; Sequence 78468, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 78468
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-78468

Query Match      21.7%; Score 10; DB 5; Length 399;
Best Local Similarity 13.3%; Pred. No. 4.1e-155;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXX 18
|
Db 45 GAAASSASAAAAASAA 59

RESULT 8
US-09-724-676A-78469
; Sequence 78469, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 78469
; LENGTH: 373
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-78469

Query Match      21.7%; Score 10; DB 5; Length 373;
Best Local Similarity 13.3%; Pred. No. 4.1e-155;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXX 18
|
Db 45 GAAASSASAAAAASAA 59

RESULT 9
US-09-724-676A-78470
; Sequence 78470, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 78470
; LENGTH: 373
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-78470

Query Match      21.7%; Score 10; DB 5; Length 373;
Best Local Similarity 13.3%; Pred. No. 4.1e-155;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXX 18
|
Db 45 GAAASSASAAAAASAA 59

RESULT 10
US-09-724-676-78468
; Sequence 78468, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 78468
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-78468

Query Match      21.7%; Score 10; DB 5; Length 399;
Best Local Similarity 13.3%; Pred. No. 4.1e-155;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXX 18
|
Db 45 GAAASSASAAAAASAA 59

RESULT 11
US-09-724-676A-78468
; Sequence 78468, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 78468
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-78468

Query Match      21.7%; Score 10; DB 5; Length 399;
Best Local Similarity 13.3%; Pred. No. 4.1e-155;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXXX 18
|
Db 45 GAAASSASAAAAASAA 59
```

```
Db 45 GAAASSASAAAASAA 59

RESULT 12
US-10-270-333-33
; Sequence 33, Application US/10270333
; GENERAL INFORMATION:
; APPLICANT: Cravchik, Anibal
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF AS INSECTICIDAL TARGETS
; FILE REFERENCE: CL000733CON
; CURRENT APPLICATION NUMBER: US/10/270,333
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: 60/168,677
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/175,691
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/191,638
; PRIOR FILING DATE: 2000-03-23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Drosophila
US-10-270-333-33

Query Match 21.7%; Score 10; DB 6; Length 462;
Best Local Similarity 13.3%; Pred. No. 4.2e-155;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18
|
Db 37 GATAATAAATTS 51

RESULT 13
US-10-131-813A-219
; Sequence 219, Application US/10131813A
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C139
; CURRENT APPLICATION NUMBER: US/10/131,813A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19

QY 4 GXXXXXXXA 18
|
Db 1016 GATTTTAAATAA 1030

RESULT 14
US-10-131-819A-219
; Sequence 219, Application US/10131819A
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Inc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C134
; CURRENT APPLICATION NUMBER: US/10/131,819A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19

QY 4 GXXXXXXXA 18
|
Db 1016 GATTTTAAATAA 1030

Query Match 21.7%; Score 10; DB 6; Length 1076;
Best Local Similarity 13.3%; Pred. No. 4.4e-155;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18
|
Db 1016 GATTTTAAATAA 1030
```

; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 219  
; LENGTH: 1076  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-131-819A-219

Query Match 21.7%; Score 10; DB 6; Length 1076;  
Best Local Similarity 13.3%; Pred. No. 4.4e-155;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXX18  
|  
Db 1016 GATTTTAATAAATA 1030

RESULT 15  
US-10-131-823A-219  
Sequence 219, Application US/10131823A  
GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Collin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C143  
; CURRENT APPLICATION NUMBER: US/10/131.823A  
; CURRENT FILING DATE: 2002-04-24  
; PRIOR APPLICATION NUMBER: 60/049911  
; PRIOR FILING DATE: 1997-06-18  
; PRIOR APPLICATION NUMBER: 60/056974  
; PRIOR FILING DATE: 1997-08-26  
; PRIOR APPLICATION NUMBER: 60/059113  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059115  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059117  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059122  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059184  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059352  
; PRIOR FILING DATE: 1997-09-19  
; PRIOR APPLICATION NUMBER: 60/059588  
; PRIOR FILING DATE: 1997-09-19  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 219  
; LENGTH: 1076  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-131-823A-219

Query Match 21.7%; Score 10; DB 6; Length 1076;  
Best Local Similarity 13.3%; Pred. No. 4.4e-155;

Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 4 GXXXXXXX18  
|  
Db 1016 GATTTTAATAAATA 1030

Search completed: January 8, 2003, 11:12:47  
Job time : 12.8103 secs



GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model  
Run on: January 8, 2003, 10:54:49 : Search time 12,1207 Seconds  
(without alignments)  
301.394 Million cell updates/sec

Title: US-09-003-869-4  
Perfect score: 46  
Sequence: 1 XXGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 38

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 10    | 21.7        | 150    | 2 T12547 | hypothetical prote |
| 2          | 10    | 21.7        | 220    | 2 JC5954 | ribosomal protein  |
| 3          | 10    | 21.7        | 305    | 2 I57039 | genomic screen hom |
| 4          | 10    | 21.7        | 314    | 2 JC5273 | paired type homeob |
| 5          | 10    | 21.7        | 333    | 2 G96780 | unknown protein f9 |
| 6          | 10    | 21.7        | 353    | 2 S44615 | mab-5 posterior-sp |
| 7          | 10    | 21.7        | 364    | 2 I48188 | gene Nkx6.1 protei |
| 8          | 10    | 21.7        | 374    | 2 T03875 | probable homeobox  |
| 9          | 10    | 21.7        | 375    | 2 T03874 | probable homeobox  |
| 10         | 10    | 21.7        | 403    | 2 A53662 | homeotic protein H |
| 11         | 10    | 21.7        | 404    | 2 B70522 | hypothetical prote |
| 12         | 10    | 21.7        | 425    | 2 S41099 | protein kinase (EC |
| 13         | 10    | 21.7        | 443    | 1 I38239 | transcription fact |
| 14         | 10    | 21.7        | 495    | 1 S31223 | transcription fact |
| 15         | 10    | 21.7        | 640    | 2 A41726 | homeotic protein B |
| 16         | 10    | 21.7        | 642    | 2 S27806 | homeotic protein B |
| 17         | 10    | 21.7        | 883    | 2 S04722 | puff 74E protein - |
| 18         | 10    | 21.7        | 1180   | 2 S69205 | stripe a/b protein |
| 19         | 10    | 21.7        | 2359   | 2 T03094 | A-kinase anchor pr |
| 20         | 10    | 21.7        | 2639   | 2 T31328 | fibroin - Chinese  |
| 21         | 10    | 21.7        | 4776   | 2 E95206 | cell wall surface  |
| 22         | 9     | 19.6        | 128    | 2 T30714 | hypothetical prote |
| 23         | 9     | 19.6        | 163    | 2 S22630 | 19K antigen - Myco |
| 24         | 9     | 19.6        | 187    | 2 S69466 | hypothetical prote |
| 25         | 9     | 19.6        | 251    | 2 S61310 | cold shock protein |
| 26         | 9     | 19.6        | 258    | 2 C86541 | CT105 hypothetical |
| 27         | 9     | 19.6        | 258    | 2 D72082 | hypothetical prote |
| 28         | 9     | 19.6        | 302    | 2 S56751 | single stranded D  |
| 29         | 9     | 19.6        | 335    | 2 G86326 | protein F18014.7 [ |

ALIGNMENTS

RESULT 1

T12547  
Hypothetical protein DKFp586E1621.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 23-Jul-1999  
C:Accession: T12547  
R:Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, June 1999  
A:Reference number: Z17528  
A:Accession: T12547  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-150 <OTT>  
A:Cross-references: EMBL:AL080235  
A:Experimental source: adult uterus; clone DKFp586E1621  
C:Genetics:  
A>Note: DKFp586E1621.1

Query Match 21.7%; Score 10; DB 2; Length 150;  
Best Local Similarity 13.3%; Pred. No. 0;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GXXXXXXXXXXXXXXA 18  
Db 126 GTTAAAAAATAAAAAA 140

RESULT 2

JC5954  
ribosomal protein L14 - human  
C:Species: Homo sapiens (man)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 02-Sep-2000  
C:Accession: JC5954  
R:Tanaka, M.; Tanaka, T.; Harata, M.; Suzuki, T.; Mitsui, Y.  
Biochem. Biophys. Res. Commun. 243, 531-537, 1998  
A:Title: Triplet repeat-containing ribosomal protein L14 gene in immortalized human e  
A:Reference number: JC5954; MUID:98153799; PMID:9480843  
A:Accession: JC5954  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-220 <TAN>  
A:Cross-references: DDBJ:P87735; NID:g1620021; PIDN:BAAL3443.1; PID:g1620022  
C:Superfamily: rat ribosomal protein L14

Query Match 21.7%; Score 10; DB 2; Length 220;  
Best Local Similarity 13.3%; Pred. No. 0;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GXXXXXXXXXXXXXXA 18  
Db 148 GTAAAAAATAAAAAA 162

## RESULT 3

I57039  
 genomic screen homeobox protein 2 - mouse  
 C:Species: Mus sp. (mouse)  
 C>Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 24-Sep-1999  
 C:Accession: I57039  
 R:Hsieh-Li, H.M.; Witte, D.P.; Szucsik, J.C.; Weinstein, M.; Li, H.; Potter, S.S.  
 Mech. Dev. 50, 177-186, 1995  
 A:Title: Gsh-2, a murine homeobox gene expressed in the developing brain.  
 A:Reference number: I57039; MUID:95344993; PMID:7619729  
 A:Accession: I57039  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-305 <RES>  
 A:Cross-references: GB:S79041; NID:gl042008; PIDN:AAB34947.1; PID:gl042009  
 C:Genetics:  
 A:Gene: Gsh-2  
 C:Superfamily: unassigned homeobox proteins; homeobox homology  
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 04-260/Domain: homeobox homology <HOX>

Query Match 21.7% Score 10; DB 2; Length 305;  
 Best Local Similarity 13.3% Pred. No. 0;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXX XXXX 18  
 |  
 Db 145 GSAAAAA AAAAAA 159

## RESULT 4

JC5273  
 paired type homeobox protein, NBP - human  
 C:Species: Homo sapiens (man)  
 C>Date: 16-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 24-Sep-1999  
 C:Accession: JC5273  
 R:Yokoyama, M.; Nishi, Y.; Yoshii, J.; Okubo, K.; Matsubara, K.  
 DNA Res. 3, 311-320, 1996  
 A:Title: Identification and cloning of neuroblastoma-specific and nerve tissue-specific  
 A:Reference number: JC5273; MUID:97191543; PMID:9039501  
 A:Contents: neuroblastoma cell  
 A:Accession: JC5273  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-314 <YOK>  
 A:Cross-references: DDBJ:D82344; NID:gl1841337; PIDN:BAAL1555.1; PID:d1012222; PID:gl1841337  
 Comment: This protein is a transcriptional repressor involved in regulating gene expression  
 C:Superfamily: unassigned homeobox proteins; homeobox homology  
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 F:99-155/Domain: homeobox homology <HOX>

Query Match 21.7% Score 10; DB 2; Length 314;  
 Best Local Similarity 13.3% Pred. No. 0;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXX XXXX 18  
 |  
 Db 240 GAAAAA AAAAAA 254

## RESULT 5

G96780  
 unknown protein F9E10.7 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: G96780  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, S.X.; Liu, Z.A.; Luros, J.S.; Mafti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: G96780  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-333 <STO>  
 A:Cross-references: GB:AE005173; NID:96646757; PIDN:AAF21069.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F9E10.7  
 A:Map position: 1

Query Match 21.7% Score 10; DB 2; Length 333;  
 Best Local Similarity 13.3% Pred. No. 0;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXX XXXX 18  
 |  
 Db 5 GATSTSA AAAAAA 19

## RESULT 6

S44615  
 mah-5 posterior-specific homeotic protein - Caenorhabditis elegans  
 N:Alternate names: C08C3.3 protein  
 C:Species: Caenorhabditis elegans  
 C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Mar-2001  
 C:Accession: S44615; A30041  
 R:Du, Z.  
 Submitted to the EMBL Data Library, May 1993  
 A:Description: Sequence of the C. elegans cosmid C08C3.  
 A:Reference number: S44615  
 A:Accession: S44615  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-353 <DUZ>  
 A:Cross-references: EMBL:L15201; NID:g289632; PIDN:AAA27945.1; PID:g289635  
 R:Costa, M.; Weir, M.; Coulson, A.; Sulston, J.; Kenyon, C.  
 Cell 55, 747-756, 1988

A:Title: Posterior pattern formation in C. elegans involves position-specific express  
 A:Reference number: A30041; MUID:89051865; PMID:2903796  
 A:Accession: A30041  
 A:Molecule type: mRNA  
 A:Residues: 143-353 <COS>  
 A:Cross-references: GB:M22751; NID:gl56361; PIDN:AAA28106.1; PID:gl56362  
 C:Genetics:  
 A:Introns: 131/2; 201/3; 229/2; 264/1; 313/3  
 C:Superfamily: unassigned homeobox proteins; homeobox homology  
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 F:271-327/Domain: homeobox homology <HOX>

Query Match 21.7% Score 10; DB 2; Length 353;  
 Best Local Similarity 13.3% Pred. No. 0;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXX XXXX 18  
 |  
 Db 182 GTSASAS SSSAAAAA 196

## RESULT 7

I48188  
 gene Nrx6.1 protein - golden hamster  
 C:Species: Mesocricetus auratus (golden hamster)  
 C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 24-Sep-1999  
 C:Accession: I48188  
 R:Rudnick, A.; Ling, T.Y.; Odagiri, H.; Rutter, W.J.; German, M.S.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 12203-12207, 1994  
 A:Title: Pancreatic beta cells express a diverse set of homeobox genes.  
 A:Reference number: I48188; MUID:95083670; PMID:7991607  
 A:Accession: I48188



A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-364 <RES>  
A:Cross-references: EMBL:X81409; NID:g587466; PIDN:CAA57166.1; PID:g587467  
C:Genetics:  
A:Gene: NKx6.1

C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:237-293/Domain: homeobox homology <HOX>

Query Match 21.7%; Score 10; DB 2; Length 364;  
Best Local Similarity 13.3%; Pred. No. 0;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXSASATSA 135

Db 121 CSSSSSSSSASATSA 135

ULT 8  
375

A:Probable homeobox protein OSH45, splice form OSH45 [similarity] - rice  
C:Species: Oryza sativa (rice)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 21-Jul-2000  
C:Accession: T03875  
R:Tamaoki, M.; Tsugawa, H.; Minami, E.; Kayano, T.; Yamamoto, N.; Kano-Murakami, Y.; Mat  
Plant J. 7, 927-938, 1995  
A:Title: Alternative RNA products from a rice homeobox gene.  
A:Reference number: 215126; MUID:95322999; PMID:7599652  
A:Accession: T03875

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-374 <TAM>

A:Cross-references: EMBL:D49704; NID:g1805615; PIDN:BAA08553.1; PID:g1805618

A:Experimental source: cv. Nipponbare

A:Comment: For alternative splice forms, see PIR:T03874.

C:Genetics:  
A:Gene: H45

A:Introns: 174/3; 215/1; 277/3; 321/3; 367/2

C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: alternative initiators; alternative splicing; DNA binding; homeobox; nucleus  
F:291-332/Domain: homeobox homology <HOX>

Query Match 21.7%; Score 10; DB 2; Length 374;  
Best Local Similarity 13.3%; Pred. No. 0;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

4 GXXXXXXXSASATSA 18

63 GAAAAAASASATSA 77

RESULT 9  
T03874

probable homeobox protein OSH45, splice form OSH44 [similarity] - rice  
N:Contains: probable homeobox protein OSH45, splice form OSH44  
C:Species: Oryza sativa (rice)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 21-Jul-2000  
C:Accession: T03874; T03876  
R:Tamaoki, M.; Tsugawa, H.; Minami, E.; Kayano, T.; Yamamoto, N.; Kano-Murakami, Y.; Mat  
Plant J. 7, 927-938, 1995  
A:Title: Alternative RNA products from a rice homeobox gene.  
A:Reference number: 215126; MUID:95322999; PMID:7599652  
A:Accession: T03874

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-375 <TAM1>

A:Cross-references: EMBL:D49704; NID:g1805615; PIDN:BAA08552.1; PID:g1805617

A:Experimental source: cv. Nipponbare, splice form OSH44

A:Accession: T03876

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 197-367 <TAM2>

A:Cross-references: EMBL:D49704; NID:g1805615; PIDN:BAA08554.1; PID:g1805619  
A:Experimental source: cv. Nipponbare, splice form OSH42  
C:Comment: For an alternative splice form, see PIR:T03875.  
C:Genetics:  
A:Gene: H45

A:Introns: 174/3; 215/1; 277/3; 321/3; 367/2

C:Superfamily: unassigned homeobox proteins; homeobox homology

C:Keywords: alternative initiators; alternative splicing; DNA binding; homeobox; nucl  
F:1-375/Product: probable homeobox protein OSH45, splice form OSH44 #status predicted  
F:197-367/Product: probable homeobox protein OSH45, splice form OSH42 #status predict  
F:291-352/Domain: homeobox homology <HOX>

Query Match 21.7%; Score 10; DB 2; Length 375;  
Best Local Similarity 13.3%; Pred. No. 0;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXSASATSA 18

Db 63 GAAAAAASASATSA 77

RESULT 10

A53662

homeotic protein HB9 - human

C:Species: Homo sapiens (man)

C:Date: 08-Jul-1995 #sequence\_revision 03-Aug-1995 #text\_change 17-Oct-1997

C:Accession: A53662

R:Harrison, K.A.; Druey, K.M.; Deguchi, Y.; Tusciano, J.M.; Kehrl, J.H.

J. Biol. Chem. 269, 19968-19975, 1994

A:Title: A novel human homeobox gene distantly related to proboscipedia is expressed

A:Reference number: A53662; MUID:94327547; PMID:7914194

A:Accession: A53662

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-403 <HAR>

A:Cross-references: GB:U07663

A:Note: the nucleotide sequence and conceptual translation as given are self-consiste

C:Genetics:  
A:Gene: GDB:HLXB9

A:Cross-references: GDB:136411; OMIM:142994

A:Map position: lq41-lq42.1

C:Superfamily: unassigned homeobox proteins; homeobox homology

C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

F:244-300/Domain: homeobox homology <HOX>

Query Match 21.7%; Score 10; DB 2; Length 403;  
Best Local Similarity 13.3%; Pred. No. 0;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXSASATSA 18

Db 120 GAAAAAASASATSA 134

RESULT 11  
B70522

hypothetical protein Rv3822 - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C:Accession: B70522

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: B70522

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-404 <COL>

A:Cross-references: GB:Z97188; GB:AL123456; NID:g3261805; PIDN:CAB10006.1; PID:e32434  
A:Experimental source: strain H37Rv

C:Genetics:  
A:Gene: RV3822

Query Match 21.7%; Score 10; DB 2; Length 404;  
Best Local Similarity 13.3%; Pred. No. 0;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXX 18  
|  
Db 360 GASAATSTAAATAAA 374

## RESULT 12

S41099  
protein kinase (EC 2.7.1.37), CAMP-dependent, catalytic chain C - fungus (Blastocladiella)  
C:Species: Blastocladiella emersonii  
C:Date: 19-Mar-1997 #sequence\_revision 06-Jun-1997 #text\_change 28-May-1999  
C:Accession: S41099; S77889; S77890  
R:Francisco de Oliveira, J.C.; Cantisani Borges, A.C.; do Valle Marques, M.; Lopes Gomes, S.  
T. J. Biochem. 219, 555-562, 1994  
A:Title: Cloning and characterization of the gene for the catalytic subunit of CAMP-depe  
Reference number: S41099; MUID:94139736; PMID:8307021

A:Accession: S41099  
A:Molecule type: DNA  
A:Residues: 1-425 <FRA>  
A:Cross-references: GB:L17008; NID:g304272; PIDN:AAA20074.1; PID:g304273  
A:Accession: S77889  
A:Molecule type: mRNA  
A:Residues: 22-425 <FRB>  
A:Cross-references: GB:M81709; GB:L17038; NID:g507140; PIDN:AAA19440.1; PID:g507141  
A:Accession: S77890  
A:Molecule type: protein  
A:Residues: 2-16 <FRC>

C:Genetics:  
A:Introns: 209/3; 243/3; 315/1  
C:Superfamily: kinase-related transforming protein; protein kinase homology  
C:Keywords: ATP; CAMP binding; magnesium; phosphoprotein; phosphotransferase; serine/thr  
F:2-425/Product: protein kinase, CAMP-dependent, catalytic chain C #status experimental  
F:114-370/Domain: protein kinase homology <KIN>  
F:122-130/Region: protein kinase ATP-binding motif  
F:127,128,194,200,243,256/Binding site: Mg-ATP (Phe, Gly, Glu, Thr) #status pr  
F:145,164,239,241/Active site: Lys, Glu, Asp, Lys #status predicted  
F:244,257/Binding site: magnesium (Asn, Asp) #status predicted  
F:270/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 21.7%; Score 10; DB 2; Length 425;  
Best Local Similarity 13.3%; Pred. No. 0;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXX 18  
|  
Db 38 GSASSTASSTTTAAA 52

## RESULT 13

I38239  
transcription factor SOX3 - human  
N:Alternate names: SRY (sex determining region Y)-box 3  
C:Species: Homo sapiens (man)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 28-Jan-2000  
C:Accession: I38239; I38242; S67816  
R:Stevanovic, M.; Lovell-Badge, R.; Collignon, J.; Goodfellow, P.N.  
Hum. Mol. Genet. 2, 2013-2018, 1993  
A:Title: SOX3 is an X-linked gene related to SRY.  
A:Reference number: I38239; MUID:94154672; PMID:8111369  
A:Accession: I38239  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-443 <STEL>  
A:Cross-references: EMBL:X71135; NID:g468790; PIDN:CAA50465.1; PID:g530020  
A:Accession: I38242  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA

A:Residues: 143-158, 'P', 160-218 <STE2>  
A:Cross-references: EMBL:X71137; NID:g468793; PIDN:CAA50467.1; PID:g468794  
C:Genetics:

A:Gene: GDB:SOX3; SOX-3; SOXB  
A:Cross-references: GDB:250376; OMIM:313430  
A:Map position: Xq26-Xq27  
C:Superfamily: human SOX3 protein; HMG box homology  
F:136-211/Domain: HMG box homology <HMG>

Query Match 21.7%; Score 10; DB 1; Length 443;  
Best Local Similarity 13.3%; Pred. No. 0;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXX 18  
|  
Db 233 GAAAAAATAAAAAA 247

## RESULT 14

S31223  
transcription factor Brn-1 - mouse  
N:Alternate names: class III POU domain protein brain-1  
C:Species: Mus musculus (house mouse)  
C:Date: 02-Dec-1993 #sequence\_revision 01-Sep-1995 #text\_change 22-Jun-1999  
C:Accession: S31223  
R:Hara, Y.; Rovescalli, A.C.; Kim, Y.; Nirenberg, M.  
Proc. Natl. Acad. Sci. U.S.A. 89, 3280-3284, 1992  
A:Title: Structure and evolution of four POU domain genes expressed in mouse brain.  
A:Reference number: S31223; MUID:92228768; PMID:1565620  
A:Accession: S31223  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-495 <HAR>

A:Cross-references: EMBL:M88299; NID:g200444; PIDN:AAA39960.1; PID:g200445  
C:Superfamily: transcription factor Brn-1; homeobox homology; POU domain homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:27-49/Region: glycine-rich  
F:101-112/Region: alanine-rich  
F:162-180/Region: histidine/proline-rich  
F:186-201/Region: alanine-rich  
F:236-247/Region: glycine-rich  
F:267-291/Region: histidine/proline-rich  
F:316-383/Domain: POU domain homology <POU>  
F:402-458/Domain: homeobox homology <HOX>

Query Match 21.7%; Score 10; DB 1; Length 495;  
Best Local Similarity 13.3%; Pred. No. 0;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXX 18  
|  
Db 185 GAAAAAATAAAAAA 199

## RESULT 15

A41726  
homeotic protein BarH2 - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 24-Sep-1999  
C:Accession: A41726  
R:Higashijima, S.; Kojima, T.; Michiue, T.; Ishimaru, S.; Emori, Y.; Saigo, K.  
Genes Dev. 6, 50-60, 1992  
A:Title: Dual Bar homeo box genes of Drosophila required in two photoreceptor cells,  
A:Reference number: A41726; MUID:92112035; PMID:1346120  
A:Accession: A41726  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-640 <HIG>  
A:Cross-references: GB:M82885; GB:M82886; GB:M82887; NID:g156986; PIDN:AAB59218.1; PI  
A>Note: the authors translated the codon TAC for residue 134 as Thr  
C:Genetics:  
A:Gene: FlyBase:B-H2  
A:Cross-references: FlyBase:FBgn0004854

C; Superfamily: unassigned homeobox proteins; homeobox homology  
C; Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F; 376-432/Domain: homeobox homology <HOX>

Query Match 21.7%; Score 10; DB 2; Length 640;  
Best Local Similarity 13.3%; Pred. No. 0;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GXXXXXXXXXXXXA 18

Db 67 GAAAAA 81

Search completed: January 8, 2003, 10:55:40  
Job time : 14.1207 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2003, 10:54:49 ; Search time 6.55172 Seconds  
(without alignments)  
240.563 Million cell updates/sec

Title: US-09-003-869-4

Perfect score: 46

Sequence: 1 XXXGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 38

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

1 number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query | Length | DB ID | Description |
|------------|-------|-------|--------|-------|-------------|
| 1          | 10    | 21.7  | 200    | 1     | MAB5_CAEEL  |
| 2          | 10    | 21.7  | 304    | 1     | GSH2_HUMAN  |
| 3          | 10    | 21.7  | 305    | 1     | GSH2_MOUSE  |
| 4          | 10    | 21.7  | 314    | 1     | PKXB_HUMAN  |
| 5          | 10    | 21.7  | 314    | 1     | PKXB_MOUSE  |
| 6          | 10    | 21.7  | 364    | 1     | HK61_MESAU  |
| 7          | 10    | 21.7  | 365    | 1     | HK61_RAT    |
| 8          | 10    | 21.7  | 367    | 1     | HK61_HUMAN  |
| 9          | 10    | 21.7  | 375    | 1     | SOX3_MOUSE  |
| 10         | 10    | 21.7  | 386    | 1     | HXAD_MOUSE  |
| 11         | 10    | 21.7  | 388    | 1     | HXAD_HUMAN  |
| 12         | 10    | 21.7  | 401    | 1     | HB9_HUMAN   |
| 13         | 10    | 21.7  | 443    | 1     | SOX3_HUMAN  |
| 14         | 10    | 21.7  | 495    | 1     | BRN1_MOUSE  |
| 15         | 10    | 21.7  | 497    | 1     | BRN1_RAT    |
| 16         | 10    | 21.7  | 500    | 1     | BRN1_HUMAN  |
| 17         | 10    | 21.7  | 563    | 1     | AXK_MOUSE   |
| 18         | 10    | 21.7  | 883    | 1     | ETAB_DROME  |
| 19         | 9     | 19.6  | 125    | 1     | PER_DROAN   |
| 20         | 9     | 19.6  | 251    | 1     | TIR2_FEAST  |
| 21         | 9     | 19.6  | 335    | 1     | HXDD_MOUSE  |
| 22         | 9     | 19.6  | 339    | 1     | HXDD_MOUSE  |
| 23         | 9     | 19.6  | 362    | 1     | NK2C_MOUSE  |
| 24         | 9     | 19.6  | 439    | 1     | ACMA_LACLA  |
| 25         | 9     | 19.6  | 508    | 1     | BTB2_MOUSE  |
| 26         | 9     | 19.6  | 525    | 1     | BTB2_HUMAN  |
| 27         | 9     | 19.6  | 548    | 1     | HEN1_YEAST  |
| 28         | 9     | 19.6  | 577    | 1     | T2FA_DROME  |
| 29         | 9     | 19.6  | 588    | 1     | CAR7_CANAL  |
| 30         | 9     | 19.6  | 609    | 1     | OPA_DROME   |
| 31         | 9     | 19.6  | 701    | 1     | TBX2_MOUSE  |
| 32         | 9     | 19.6  | 702    | 1     | TBX2_HUMAN  |
| 33         | 9     | 19.6  | 829    | 1     | E74A_DROME  |

34 9 19.6 851 1 NCL1\_CAEEL P34611 caenorhabdi  
35 9 19.6 1205 1 NKCI\_MOUSE P55012 mus musculu  
36 9 19.6 1211 1 BUN2\_DROME Q34523 drosophila  
37 9 19.6 1212 1 NKCI\_HUMAN P55011 homo sapien  
38 9 19.6 1533 1 PUM\_DROME P35822 drosophila  
39 9 19.6 1944 1 CHD3\_HUMAN Q12873 homo sapien  
40 9 19.6 2038 1 FSH\_DROME P3709 drosophila  
41 8 17.4 105 1 RLA2\_LEIDO O43940 teishmania  
42 8 17.4 132 1 HEX9\_ADE40 P48312 human adeno  
43 8 17.4 133 1 HEX9\_ADE41 P32539 human adeno  
44 8 17.4 173 1 CH19\_DROME P07186 drosophila  
45 8 17.4 208 1 PAL1\_CAEEL P34766 caenorhabdi

#### ALIGNMENTS

RESULT 1  
MAB5\_CAEEL  
ID MAB5\_CAEEL STANDARD: PRT: 200 AA.  
AC P10038: Q9GZ02:  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Homeobox protein mab-5 (Male abnormal-5 protein).  
GN MAB-5 OR C08C3.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89051865; PubMed=2903796;  
RA Costa M., Wir M., Coulson A., Sulston J., Kenyon C.;  
RT "Posterior pattern formation in C. elegans involves position-specific  
RT expression of a gene containing a homeobox.";  
RL Cell 55:747-756(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Grandien K., Sommer R.J.;  
RT "Determination of 5' end of Caenorhabditis elegans mab-5 cDNA  
RT demonstrates a shorter N-terminal region than previously predicted.";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Durbin R., Favello A., Fraser A.,  
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,  
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,  
RA Sulston J., Thierry-Mieg J., Thomas K., Yaudin M., Vaughan K.,  
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
RN [4]  
RP REVISIONS.  
RA Waterston R.;  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: DURING POSTEMBRYONIC DEVELOPMENT IN C.ELEGANS,  
CC POSTERIOR-SPECIFIC PATTERN FORMATION REQUIRES MAB-5. WITHIN THE  
CC POSTERIOR BODY REGION, MAB-5 ACTIVITY CONTROLS EPIDERMAL,  
CC NEURONAL, AND MESODERMAL CELL DIFFERENTIATION.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- SIMILARITY: BELONGS TO THE ANTP HOMEBOX FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M22751; AAA28106.1; ALT\_INIT.  
 DR EMBL; AF277990; AAG00458.1; -.  
 DR EMBL; L15201; AAA27945.2; -.  
 DR PIR; A30041; A30041.  
 DR PIR; S44615; S44615.  
 DR HSP; P02633; 9ANT.  
 DR TRANSFAC; T03368; -.  
 DR WormPeP; C08C3.3; CE25765.  
 DR InterPro; IPR001827; Antennapedia.  
 DR InterPro; IPR001356; Homeobox.  
 DR Pfam; PF00046; homeobox; 1.  
 DR PRINTS; PR00025; ANTENNAPEIDIA.  
 DR PRINTS; PR00024; HOMEBOX.  
 DR ProDom; PD000010; Homeobox; 1.  
 DR SMART; SM00389; HOX; 1.  
 DR PROSITE; PS00027; HOMEBOX\_1; 1.  
 DR PROSITE; PS50071; HOMEBOX\_2; 1.  
 DR PROSITE; PS00032; ANTENNAPEIDIA; 1.  
 DR Homeobox; DNA-binding; Nuclear protein; Developmental protein.  
 DR DOMAIN 19 44 ALA/THR/SER-RICH.  
 DR DOMAIN 38 44 POLY-ALA.  
 DR DOMAIN 103 108 ANTP-TYPE HEXAPEPTIDE.  
 DR DOMAIN 115 175 HOMEBOX.  
 DR DNA\_BIND 115 175 HOMEBOX.  
 DR SEQUENCE 200 AA; 22398 MW; EAC2DCA86F54E7E3 CRC64;

Query Match 21.7%; Score 10; DB 1; Length 200;  
 Best Local Similarity 13.3%; Pred. No. 4.4e-79;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXX 18  
 DB 29 GTSASASSSSAAAAA 43

RESULT 2  
 GSH2\_HUMAN STANDARD; PRT; 304 AA.  
 ID GSH2\_HUMAN  
 AC Q9BZM3;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Homeobox protein GSH-2.  
 GN GSH2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sakai T., Sakamoto S., Nakamura K., Muraki T.;  
 RT "Human homeobox protein GSH-2";  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Coles J., Marynen P.;  
 RT "The sequence of the human GSH2 gene";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR THAT BINDS TO THE DNA  
 CC SEQUENCE 5'-CNAATTAG-3' (by similarity).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (by similarity).  
 CC -!- SIMILARITY: BELONGS TO THE ANTP HOMEBOX FAMILY.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AB028838; BAB84822.1; -.  
 DR EMBL; AF306344; AAK00880.1; -.  
 DR EMBL; AF306343; AAK00880.1; JOINED.  
 DR HSP; P14653; 1B72.  
 DR InterPro; IPR000047; HTH\_repressr.  
 DR InterPro; IPR001356; Homeobox.  
 DR Pfam; PF00046; homeobox; 1.  
 DR PRINTS; PR00024; HOMEBOX.  
 DR PRINTS; PR00031; HTHREPRESSR.  
 DR ProDom; PD000010; Homeobox; 1.  
 DR SMART; SM00389; HOX; 1.  
 DR PROSITE; PS00027; HOMEBOX\_1; 1.  
 DR PROSITE; PS50071; HOMEBOX\_2; 1.  
 DR Transcription regulation; Homeobox; DNA-binding; Nuclear protein;  
 KW Developmental protein.  
 FT DNA\_BIND 202 261 HOMEBOX.  
 FT DOMAIN 124 130 POLY-HIS.  
 FT DOMAIN 134 139 POLY-HIS.  
 FT DOMAIN 147 162 POLY-ALA.  
 FT SEQUENCE 304 AA; 32061 MW; E896D5422488E6C1 CRC64;

Query Match 21.7%; Score 10; DB 1; Length 304;  
 Best Local Similarity 13.3%; Pred. No. 4.5e-79;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXX 18  
 DB 145 GSAAAAAASAAAAA 159

RESULT 3  
 GSH2\_MOUSE STANDARD; PRT; 305 AA.  
 ID GSH2\_MOUSE  
 AC P31316;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Homeobox protein GSH-2.  
 GN GSH2 OR GSH-2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NIH Swiss;  
 RX MEDLINE=95344993; PubMed=7619729;  
 RA Hsieh-Li A.M., Witte D.P., Szucsik J.C., Weinstein M., Li H.,  
 RT "Gsh-2, a murine homeobox gene expressed in the developing brain";  
 RL Mech. Dev. 50:177-186(1995).  
 RN [2]  
 RP SEQUENCE OF 203-262 FROM N.A.  
 RX MEDLINE=92073356; PubMed=1683707;  
 RA Singh G., Kaur S., Stock J.L., Jenkins N.A., Gilbert D.J.,  
 RA Copeland N.G., Potter S.S.;  
 RT "Identification of 10 murine homeobox genes";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10706-10710(1991).  
 CC -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR THAT BINDS TO THE DNA  
 CC SEQUENCE 5'-CNAATTAG-3'.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE DEVELOPING BRAIN.  
 CC -!- SIMILARITY: BELONGS TO THE ANTP HOMEBOX FAMILY.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

|    |   |
|----|---|
| CC | EMBL: S79041; AAB34947.1; -                                       |
| DR | PIR: B37290; B37290.  |
| DR | PIR: B38809; B38809.  |
| DR | HSSP: P14653; 1B72.   |
| DR | MGO: MGI:95843; Gsh2.   |
| DR | InterPro: IPR001356; Homeobox.                                    |
| DR | Pfam: PF00046; homeobox; 1.                                       |
| DR | PRINTS: PR00024; HOMEBOX.   |
| DR | ProDom: PD000010; Homeobox; 1.                                    |
| DR | SMART: SM00389; HOX; 1.   |
| DR | PROSITE: PS00027; HOMEBOX_1; 1.                                   |
| DR | PROSITE: PS00071; HOMEBOX_2; 1.                                   |
| DR | Transcription regulation; Homeobox; DNA-binding; Nuclear protein; |
| KW | Developmental protein.  |
| FT | DNA_BIND 203 262 HOMEBOX.   |
| FT | DOMAIN 124 130 POLY-HIS.  |
| FT | DOMAIN 134 139 POLY-HIS.  |
| FT | DOMAIN 147 163 POLY-ALA.  |
| FT | SEQUENCE 305 AA; 32167 MW; 51E7F2DB76E32608 CRC64;                |

```

Query Match          21.7%; Score 10; DB 1; Length 305;
Best Local Similarity 13.3%; Pred. No. 4.5e-79;
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY      4 GXXXXXXXXXXXXX 18
      |
Db      145 GAAAAAAAAAAAAA 159

RESULT 4
ID      PMXB_HUMAN      STANDARD;      PRT;      314 AA.
AC      Q99453;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Paired mesoderm homeobox protein 2B (Paired-like homeobox 2B)
DE      (PHOX2B homeodomain protein) (Neuroblastoma Phox) (NRPHox).
DE      PMXB2B.
GN      CS      Homo sapiens (Human).
OS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
EN      11

```

CC PHENOTYPE IN VERTEBRATES. ENHANCES SECOND-MESSENGER-MEDIATED  
CC ACTIVATION OF THE DOPAMINE BETA-HYDROXYLASE AND C-FOS PROMOTERS,  
CC AND OF SEVERAL ENHANCERS INCLUDING CYCLIC AMP-RESPONSE ELEMENT AND  
CC SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN NEUROBLASTOMA, BRAIN AND ADRENAL  
CC GLAND.  
CC -!- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: D82344; BAAL1555.1; --  
CC EMBL: AF117979; AAD26698.1; --  
CC EMBL: AB015671; BAA82670.1; --  
CC HSSP: P06601; 1FJL.  
CC TRANSFAC: T03961; --  
CC GeneW: HGNC:9143; PMX2B.  
CC MIM: 603851; --  
CC InterPro: IPR001356; Homeobox.  
CC Pfam: PF000046; homeobox; 1.  
CC PRINTS: PR00024; HOMEBOX.  
CC ProDom: PD000010; Homeobox; 1.  
CC SMART: SM00389; HOX\_1.  
CC PROSITE: PS00027; HOMEBOX\_1; 1.  
CC PROSITE: PSS0071; HOMEBOX\_2; 1.  
CC Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
CC Transcription regulation.  
CC DNA\_BIND 98 157 HOMEBOX.  
CC DOMAIN 159 167 POLY-ALA.  
CC DOMAIN 212 217 POLY-GLY.  
CC DOMAIN 241 260 POLY-ALA.  
CC SEQUENCE 314 AA; 31607 MW; 76737F71948B5D81 CRC64;  
CC  
CC Query Match 21.7%; Score 10; DB 1; Length 314;  
CC Best Local Similarity 13.3%; Pred. No. 4.5e-79;  
CC Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
CC  
CC QY 4 GXNXXXXXXXXXXA 18  
CC |  
CC Db 240 GAAAAAAGAAAAA 254  
CC  
CC RESULT 5  
CC PMXB\_MOUSE  
CC ID PMXB\_MOUSE STANDARD; PRT; 314 AA.  
CC OS O35690;  
CC DT 15-JUL-1999 (Rel. 38, Created)  
CC DT 15-JUL-1999 (Rel. 38, Last sequence update)  
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
CC DE Paired mesoderm homeobox protein 2B (paired-like homeobox 2B)  
CC (PHOX2B homeodomain protein) (Neuroblastoma Phox) (NBphox).  
CC GN PMXB2B OR PHOX2B.  
CC OS Mus musculus (Mouse).  
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.  
CC RN NCBI\_TaxId=10090;  
CC RX [1]  
CC SEQUENCE FROM N.A.  
CC RX MEDLINE=98040559; PubMed=9374403;  
CC RX Patten A., Morin X., Cremer H., Goridis C., Brunet J.-F.;  
CC RT "Expression and interactions of the two closely related homeobox  
CC RL genes Phox2a and Phox2b during neurogenesis."; Development 124:4065-4075(1997).  
CC RX [2]  
CC SEQUENCE FROM N.A.  
CC RX MEDLINE=99326521; PubMed=10395798;  
CC RX Yokoyama M., Watanabe H., Nakamura M.;

RT "Genomic structure and functional characterization of NBphox (PMX2B),  
 RT a homeodomain protein specific to catecholaminergic cells that is  
 RT involved in second messenger-mediated transcriptional activation.;  
 RL Genomics 59:40-50(1999).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: Y14493; CAA74833.1; -.  
 CC EMBL: AB015672; BAA82671.1; -.  
 CC HSSP: P06601; IFJL.  
 CC TRANSFAC: T03976; -.  
 CC MGD: MGI:1100682; Pmx2b.  
 CC InterPro: IPR000047; HTH\_repressr.  
 CC Pfam: PF00046; Homeobox; 1.  
 CC PRINTS: PR00024; HOMEBOX.  
 CC PRINTS: PR00031; HTHREPRESSR.  
 CC ProDom: PD000010; Homeobox; 1.  
 CC SMART: SM00389; HOX; 1.  
 CC PROSITE: PS00027; HOMEBOX\_1; 1.  
 CC PROSITE: PS00071; HOMEBOX\_2; 1.  
 CC Homeobox; DNA-binding; Developmental protein; Nuclear protein.  
 KW DOMAIN 98 157 HOMEBOX.  
 FT DNA\_BIND 98 157 HOMEBOX.  
 FT DOMAIN 159 167 POLY-ALA.  
 FT DOMAIN 212 217 POLY-GLY.  
 FT DOMAIN 241 250 POLY-ALA.  
 FT SEQUENCE 314 AA; 31621 MR; 40737F71948B595A CRC64;  
 SQ  
 Query Match 21.7%; Score 10; DB 1; Length 314;  
 Best Local Similarity 13.3%; Pred. No. 4.5e-79;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 QY 4 GXXXXXXXXXXXX 18  
 Db 240 GAAAAAAAAAAAAA 254  
 RESULT 6  
 HK61\_MESAU STANDARD; PRT; 364 AA.  
 Q60554;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Homeobox protein NKX-6.1.  
 GN NKX6A Or NKX6.1.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 CC Mesocricetus.  
 CC NCBI\_TaxID=10036;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreatic islets;  
 RX MEDLINE=95083670; PubMed=7991607;  
 RA Rudnick A., Ling T.Y., Odagiri H., Rutter W.J., German M.S.;  
 RT "Pancratic beta cells express a diverse set of homeobox genes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:12203-12207(1994).  
 CC -!- FUNCTION: MAY BE IMPORTANT FOR CONTROL OF ISLET DEVELOPMENT AND/OR  
 CC REGULATION OF INSULIN BIOSYNTHESIS.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -!- TISSUE SPECIFICITY: PANCREATIC BETA CELLS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X81409; CAA57166.1; -.  
 CC HSSP: P06601; IFJL.  
 CC TRANSFAC: T04261; -.  
 CC InterPro: IPR000047; HTH\_repressr.  
 CC Pfam: PF00046; Homeobox; 1.  
 CC PRINTS: PR00024; HOMEBOX.  
 CC PRINTS: PR00031; HTHREPRESSR.  
 CC ProDom: PD000010; Homeobox; 1.  
 CC SMART: SM00389; HOX; 1.  
 CC PROSITE: PS00027; HOMEBOX\_1; 1.  
 CC PROSITE: PS00071; HOMEBOX\_2; 1.  
 CC Homeobox; DNA-binding; Developmental protein; Nuclear protein.  
 KW DOMAIN 49 60 POLY-SER.  
 FT DOMAIN 118 131 POLY-SER.  
 FT DOMAIN 135 150 POLY-ALA.  
 FT DOMAIN 168 173 POLY-PRO.  
 FT DNA\_BIND 236 295 HOMEBOX.  
 FT SEQUENCE 364 AA; 37646 MW; 08B1AE1225F0F06E CRC64;  
 SQ  
 Query Match 21.7%; Score 10; DB 1; Length 364;  
 Best Local Similarity 13.3%; Pred. No. 4.6e-79;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 QY 4 GXXXXXXXXXXXX 18  
 Db 121 GSSSSSSSSSASATSA 135  
 RESULT 7  
 HK61\_RAT STANDARD; PRT; 365 AA.  
 AC Q35762;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Homeobox protein NKX-6.1.  
 GN NKX6A Or NKX6.1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wistar; TISSUE=Pancreatic islets;  
 RX MEDLINE=20036461; PubMed=10567713;  
 RA Jorgensen M.C., Vestergaard Petersen H., Ericson J., Madsen O.D.,  
 RA Serup P.;  
 RT "Cloning and DNA-binding properties of the rat pancreatic  
 RT beta cell-specific factor NKX6.1.";  
 RL FEBS Lett. 461:287-294(1999).  
 CC -!- FUNCTION: MAY BE IMPORTANT FOR CONTROL OF ISLET DEVELOPMENT AND/OR  
 CC REGULATION OF INSULIN BIOSYNTHESIS (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -!- TISSUE SPECIFICITY: PANCREATIC BETA CELLS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF004431; AAB61665.1; -.  
 CC HSSP: P06601; IFJL.  
 CC TRANSFAC: T04297; -.





DR SMART: SM00398; HMG; 1.  
 KW DNA-binding; Nuclear protein; Transcription regulation.  
 FT DOMAIN 54 63 HMG-GLY.  
 FT DNA\_BIND 69 137 HMG BOX.  
 FT DOMAIN 164 177 POLY-ALA.  
 FT DOMAIN 219 223 POLY-PRO.  
 FT DOMAIN 252 259 POLY-ALA.  
 FT DOMAIN 269 276 POLY-ALA.  
 FT DOMAIN 282 293 POLY-ALA.  
 SQ SEQUENCE 375 AA; 37857 MW; 2A8477A10A517FEF CRC64;

Query Match 21.7%; Score 10; DB 1; Length 375;  
 Best Local Similarity 13.3%; Pred. No. 4.6e-79;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXX 18  
 |  
 DB 163 GAAAAAAXXXXXX 177

RESULT 10  
 ID HXAD\_MOUSE STANDARD; PRT; 386 AA.  
 AC Q62424;

DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Homeobox protein Hox-A13 (Hox-1.10).

GN HOXA13 OR HOX-1.10.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96259555; PubMed=8673126;

RA Mortlock D.P., Post L.C., Innis J.W.;

RT "The molecular basis of hypodactyly (HD): a deletion in Hoxa 13 leads

to arrest of digital arch formation.";

RL Nat. Genet. 13:284-289(1996).

CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF

A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH

SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS (BY

SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- DISEASE: DEFECTS IN HOXA13 ARE THE CAUSE OF HYPODACTYLY (HD), A

CONDITION CHARACTERIZED BY PROFOUND DEFICIENCY OF DIGITAL ARCH

STRUCTURES.

CC -!- SIMILARITY: BELONGS TO THE ABD-B HOMEBOX FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL Outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

-----

DR EMBL: U59322; AAB03322.1;

DR HSSP: P14653; I872.

DR TRANSFAC: T03337;

DR MGD: MGI:96173; Hoxa13.

DR InterPro: IPR001356; Homeobox.

DR Pfam: PF00046; Homeobox; 1.

DR ProDom: PD000010; Homeobox; 1.

DR SMART: SM00389; HOX; 1.

DR PROSITE: PS00027; HOMEBOX\_1; 1.

DR PROSITE: PS50071; HOMEBOX\_2; 1.

DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;

Transcription regulation.

FT DOMAIN 38 51 POLY-ALA.

FT DNA\_BIND 320 379 HOMEBOX.

FT DOMAIN 52 57 POLY-GLY.

FT DOMAIN 62 66 POLY-ALA.  
 FT DOMAIN 73 84 POLY-ALA.  
 FT DOMAIN 101 104 POLY-ALA.  
 FT DOMAIN 116 133 POLY-ALA.  
 FT DOMAIN 198 205 POLY-ALA.  
 SQ SEQUENCE 386 AA; 39566 MW; 2B01DCC9B1951324 CRC64;

Query Match 21.7%; Score 10; DB 1; Length 386;  
 Best Local Similarity 13.3%; Pred. No. 4.6e-79;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXXX 18  
 |  
 DB 37 GAAAAAAXXXXXX 51

RESULT 11

ID HXAD\_HUMAN

AC P31271; O43371; STANDARD; PRT; 388 AA.

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Homeobox protein Hox-A13 (Hox-1J).

GN HOXA13 OR HOX1J.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97172976; PubMed=9020844;

RA Mortlock D.P., Innis J.W.;

RT "Mutation of HOXA13 in hand-foot-genital syndrome.";

RL Nat. Genet. 15:179-180(1997).

CC [2]

RP SEQUENCE FROM N.A.

RA Bradshaw H., Hinds K., Keppler D.;

RT Submitted (Feb-1998) to the EMBL/GenBank/DBJ databases.

RL [3]

RP SEQUENCE OF 322-387 FROM N.A.

RX MEDLINE=90098876; PubMed=2574852;

RA Acampora D., D'Esposito M., Falella A., Pannese M., Migliaccio E.,

RA Morelli F., Stornaiuolo A., Nigro V., Simeone A., Boncinelli E.;

RT "The human Hox gene family.";

RL Nucleic Acids Res. 17:10385-10402(1989).

CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF

A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH

SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- SIMILARITY: BELONGS TO THE ABD-B HOMEBOX FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL Outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

-----

DR EMBL: U82827; AAC50993.1;

DR EMBL: AC004080; NOT\_ANNOTATED\_CDS.

DR PIR: S14932; S14932.

DR HSSP: P14653; I872.

DR TRANSFAC: T03321;

DR GENE: HGNC:5102; HOXA13.

DR MIM: 142959;

DR InterPro: IPR001356; Homeobox.

DR Pfam: PF00046; homeobox; 1.

DR ProDom: PD000010; Homeobox; 1.

DR SMART: SM00389; HOX; 1.

DR PROSITE: PS00027; HOMEBOX\_1; 1.

DR PROSITE: PS50071; HOMEBOX\_2; 1.

KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
Transcription regulation.

FT DNA\_BIND 322 381 HOMEBOX.  
FT DOMAIN 38 53 POLY-ALA.  
FT DOMAIN 62 66 POLY-ALA.  
FT DOMAIN 73 84 POLY-ALA.  
FT DOMAIN 116 133 POLY-ALA.  
FT DOMAIN 145 150 POLY-ALA.  
FT DOMAIN 200 207 POLY-ALA.  
FT CONFLICT 146 146 A -> G (IN REF. 2).  
FT CONFLICT 187 187 P -> H (IN REF. 2).  
FT CONFLICT 195 195 P -> A (IN REF. 2).  
FT CONFLICT 198 198 P -> A (IN REF. 2).  
SQ SEQUENCE 388 AA: 39752 MW: 6CD9CA5616C2FF6 CRC64;

Query Match 21.7%; Score 10; DB 1; Length 388;  
Best Local Similarity 13.3%; Pred. No. 4.6e-79;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

4 GXXXXXXXXXXXXXA 18  
|  
37 GAAAAAAAAAAAAA 51

#### RESULT 12

HB9\_HUMAN  
ID HB9\_HUMAN STANDARD; PRT; 401 AA.  
AC P50219;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Homeobox protein HB9.  
GN HLXB9.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94327547; PubMed=7914194;  
RA Harrison K.A., Druey K.M., Deguchi Y., Tusciano J.M., Kehrl J.H.;  
RT "A novel human homeobox gene distantly related to proboscipedia is  
expressed in lymphoid and pancreatic tissues.";  
RL J. Biol. Chem. 269:19968-19975(1994).  
CC -1- FUNCTION: PUTATIVE TRANSCRIPTION FACTOR.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN LYMPHOID AND PANCREATIC TISSUES.  
CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; U07664; AAB60647.1; -.  
DR EMBL; U07663; AAB60647.1; JOINED.  
DR HSSP; P14653; 1B72.  
DR TRANSFAC; T03420; -.  
DR Genew; HGNC:4979; HLXB9.  
DR MIM; 142994; -.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox; 1.  
DR PRINTS; PR00024; HOMEBOX.  
DR ProDom; PD000010; Homeobox; 1.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
KW Homeobox; DNA-binding; Homeobox; Nuclear protein; Transcription regulation.  
FT DOMAIN 39 48 POLY-GLY.

FT DOMAIN 97 111 POLY-GLY.  
FT DOMAIN 120 135 POLY-ALA.  
FT DOMAIN 169 177 POLY-ALA.  
FT DNA\_BIND 242 301 HOMEBOX.  
FT DOMAIN 316 325 POLY-GLY.  
SQ SEQUENCE 401 AA: 40932 MW: 0006AEAD71D594FE CRC64;

Query Match 21.7%; Score 10; DB 1; Length 401;  
Best Local Similarity 13.3%; Pred. No. 4.6e-79;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 4 GXXXXXXXXXXXXXA 18  
|  
Db 119 GAAAAAAAAAAAAA 133

#### RESULT 13

SOX3\_HUMAN  
ID SOX3\_HUMAN STANDARD; PRT; 443 AA.  
AC P41225; P35714;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Transcription factor SOX-3.  
GN SOX3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94154672; PubMed=8111369;  
RA Stevanovic M., Lovell-Badge R., Collignon J., Goodfellow P.N.;  
RT "SOX3 is an X-linked gene related to SRY.";  
RL Hum. Mol. Genet. 2:2013-2018(1993).  
RN [2]  
RP SEQUENCE OF 150-203 FROM N.A.  
RX MEDLINE=92310993; PubMed=1614875;  
RA Denny P., Swift S., Brand N., Dabhadre N., Barton P., Ashworth A.;  
RT "A conserved family of genes related to the testis determining gene,  
SRY.";  
RL Nucleic Acids Res. 20:2887-2887(1992).  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.  
CC -1- CAUTION: WAS CALLED SOX-9 BY REF. 2.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; X71135; GAA50465.1; -.  
DR EMBL; Y65665; GAA46616.1; -.  
DR PIR; S21483; S21483.  
DR PIR; S22942; S22942.  
DR HSSP; Q05066; 1HRY.  
DR TRANSFAC; T04916; -.  
DR Genew; HGNC:11199; SOX3.  
DR MIM; 313430; -.  
DR InterPro; IPR000910; HMG\_12\_box.  
DR Pfam; PF00505; HMG\_box; 1.  
DR SMART; SM00398; HMG; 1.  
KW DNA-binding; Nuclear protein; Transcription regulation.  
FT DOMAIN 129 133 POLY-GLY.  
FT DNA\_BIND 139 207 HMG\_BOX.  
FT DOMAIN 234 248 POLY-ALA.  
FT DOMAIN 290 294 POLY-PRO.  
FT DOMAIN 321 327 POLY-ALA.  
FT DOMAIN 337 344 POLY-ALA.  
FT DOMAIN 350 361 POLY-ALA.

FT CONFLICT 159 159 L -> Q (IN REF. 2).  
FT CONFLICT 176 176 D -> E (IN REF. 2).  
FT CONFLICT 202 202 E -> D (IN REF. 2).  
SQ SEQUENCE 443 AA; 44884 MW; 8031B4EADA52D3B4 CRC64;

Query Match 21.7%; Score 10; DB 1; Length 443;  
Best Local Similarity 13.3%; Pred. No. 4.6e-79;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GXXXXXXXA 18  
|  
Db 233 GAAAAA 247

RESULT 14  
BRN1\_MOUSE  
ID BRN1\_MOUSE STANDARD; PRT; 495 AA.  
AC P31361;  
DT 01-JUL-1993 (Rel. 26, Created)  
15-JUL-1993 (Rel. 26, Last sequence update)  
15-JUN-2002 (Rel. 41, Last annotation update)  
Brain-specific homeobox/POU domain protein 1 (BRN-1 protein).  
POU3F3 OR OTF8 OR BRN1 OR BRN-1.

GN Mus musculus (Mouse).  
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE=92228768; PubMed=1565620;  
RA Hara Y., Rovescalli C., Kim Y., Nirenberg M.;  
RT "Structure and evolution of four POU domain genes expressed in mouse brain";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:3260-3284(1992).

CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- TISSUE SPECIFICITY: BRAIN.  
CC -1- SIMILARITY: BELONGS TO THE POU TRANSCRIPTION FACTOR FAMILY.

CC CLASS-3 SUBFAMILY.  
CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M88299; AAA39960.1; -  
PIR; S31223; S31223.  
HSSP; P14859; 10CT.  
MGD; MGI:102564; Pou3f3.  
InterPro; IPR001356; Homeobox.  
InterPro; IPR000327; POU\_domain.  
Pfam; PF00046; homeobox; 1.  
Pfam; PF00157; pou; 1.  
PRINTS; PR00028; POU\_DOMAIN.  
ProDom; PD000010; Homeobox; 1.  
ProDom; PD000583; POU\_domain; 1.  
SMART; SM00389; HOX; 1.  
SMART; SM00352; POU; 1.  
PROSITE; PS00028; POU\_DOMAIN.  
PROSITE; PS50071; HOMEBOX\_1; 1.  
PROSITE; PS50071; HOMEBOX\_2; 1.  
PROSITE; PS00035; POU\_1; 1.  
PROSITE; PS00465; POU\_2; 1.  
Nuclear protein; DNA-binding; Homeobox.  
FT DOMAIN 28 49  
FT POLY-GLY.  
FT DOMAIN 101 112  
FT POLY-ALA.  
FT DOMAIN 186 201  
FT POLY-ALA.  
FT DOMAIN 267 291  
FT HIS-RICH.  
FT DOMAIN 313 383  
FT POU.  
FT DNA\_BIND 401 460  
FT HOMEBOX.  
SQ SEQUENCE 495 AA; 50012 MW; 77B802E890C9A014 CRC64;

Query Match 21.7%; Score 10; DB 1; Length 495;  
Best Local Similarity 13.3%; Pred. No. 4.6e-79;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GXXXXXXXA 18  
|  
Db 185 GAAAAA 199

RESULT 15

BRN1\_RAT  
ID BRN1\_RAT STANDARD; PRT; 497 AA.  
AC Q63262;

DT 15-JUL-1998 (Rel. 36, Created)  
15-JUL-1999 (Rel. 38, Last sequence update)  
15-JUN-2002 (Rel. 41, Last annotation update)  
Brain-specific homeobox/POU domain protein 1 (BRN-1 protein).  
POU3F3 OR BRN1 OR BRN-1 OR RHS2.

GN Rattus norvegicus (Rat).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;

RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98070400; PubMed=9405434;  
RA Schreiber J., Enderich J., Sock E., Schmidt C., Richter-Landsberg C.,  
Wegner M.;  
RT "Redundancy of class III POU proteins in the oligodendrocyte lineage";  
RL J. Biol. Chem. 272:32286-32293(1997).

RN [2]  
RP SEQUENCE OF 325-449 FROM N.A.  
TISSUE-Hypothalamus;  
RX MEDLINE=92228769; PubMed=1348858;  
RA le Moine C., Young W.S.;

RT "RHS2, a POU domain-containing gene, and its expression in developing and adult rat";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:3285-3289(1992).

CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -1- TISSUE SPECIFICITY: BRAIN.

CC -1- DEVELOPMENTAL STAGE: EXPRESSED FROM EMBRYONIC DAY 11.5 INTO ADULTHOOD.

CC -1- SIMILARITY: BELONGS TO THE POU TRANSCRIPTION FACTOR FAMILY.

CC CLASS-3 SUBFAMILY.

CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AJ01641; CAA04893.1; -  
EMBL; M84644; AAA2041.1; -  
HSSP; P14859; 10CT.  
InterPro; IPR001356; Homeobox.  
InterPro; IPR000327; POU\_domain.  
Pfam; PF00046; homeobox; 1.  
Pfam; PF00157; pou; 1.  
PRINTS; PR00028; POU\_DOMAIN.  
ProDom; PD000010; Homeobox; 1.  
ProDom; PD000583; POU\_domain; 1.  
SMART; SM00389; HOX; 1.  
SMART; SM00352; POU; 1.  
PROSITE; PS00027; HOMEBOX\_1; 1.  
PROSITE; PS50071; HOMEBOX\_2; 1.  
PROSITE; PS00035; POU\_1; 1.  
PROSITE; PS00465; POU\_2; 1.  
Nuclear protein; DNA-binding; Homeobox.  
FT DOMAIN 28 49  
FT POLY-GLY.

FT DOMAIN 103 114 POLY-ALA.  
FT DOMAIN 135 143 POLY-PRO.  
FT DOMAIN 173 178 POLY-PRO.  
FT DOMAIN 188 203 POLY-ALA.  
FT DOMAIN 238 249 POLY-GLY.  
FT DOMAIN 269 280 POLY-HIS.  
FT DOMAIN 294 301 POLY-GLY.  
FT DOMAIN 315 385 POU.  
FT DNA\_BIND 403 462 HOMEBOX.  
SQ SEQUENCE 497 AA: 50226 MW: 00640505E343ABC2 CRC64;

Query Match 21.7%; Score 10; DB 1; Length 497;  
Best Local Similarity 13.3%; Pred. No. 4.6e-79;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXXXXA 18  
Db 187 GAAAAAAAAAAAAA 201

Search completed: January 8, 2003, 10:58:53  
Sub time : 9.55172 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2003, 10:54:50 : Search time 22.6034 Seconds  
(without alignments)  
346.398 Million cell updates/sec

Title: US-09-003-869-4  
Perfect score: 46  
Sequence: 1 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 38

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
All number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL\_21.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phase.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*
  - 15: sp\_rvirus.\*
  - 16: sp\_bacteriap.\*
  - 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID        | Description        |
|------------|-------|-------------|--------|-----------|--------------------|
| 1          | 10    | 21.7        | 130    | 6 Q9TUC9  | Q9tuc9 canis famil |
| 2          | 10    | 21.7        | 131    | 6 Q9TST8  | Q9tst8 felis silve |
| 3          | 10    | 21.7        | 150    | 4 Q9Y4M1  | Q9y4m1 homo sapien |
| 4          | 10    | 21.7        | 151    | 5 Q95VQ0  | Q95vq0 antheraea y |
| 5          | 10    | 21.7        | 191    | 4 Q96NI3  | Q96ni3 homo sapien |
| 6          | 10    | 21.7        | 193    | 11 Q9D6J0 | Q9d6j0 mus musculu |
| 7          | 10    | 21.7        | 200    | 5 Q9G202  | Q9g202 caenorhabdi |
| 8          | 10    | 21.7        | 202    | 5 Q9G203  | Q9g203 caenorhabdi |
| 9          | 10    | 21.7        | 213    | 6 Q9TUC8  | Q9tuc8 monodelphis |
| 10         | 10    | 21.7        | 218    | 4 Q96GR0  | Q96gr0 homo sapien |
| 11         | 10    | 21.7        | 219    | 4 Q8WUT0  | Q8wut0 homo sapien |
| 12         | 10    | 21.7        | 222    | 4 Q9BSB8  | Q9bsb8 homo sapien |
| 13         | 10    | 21.7        | 224    | 11 Q9ERQ7 | Q9erq7 mus musculu |
| 14         | 10    | 21.7        | 236    | 4 Q9BX46  | Q9bx46 homo sapien |
| 15         | 10    | 21.7        | 285    | 11 Q91XV7 | Q91xv7 rattus norv |
| 16         | 10    | 21.7        | 286    | 13 Q9PUX6 | Q9pux6 gadus morhu |

|    |    |      |     |           |                    |
|----|----|------|-----|-----------|--------------------|
| 17 | 10 | 21.7 | 300 | 4 Q8W271  | Q8w271 homo sapien |
| 18 | 10 | 21.7 | 327 | 5 Q9V5M5  | Q9v5m5 drosophila  |
| 19 | 10 | 21.7 | 333 | 10 Q9C9Q2 | Q9c9q2 arabidopsis |
| 20 | 10 | 21.7 | 336 | 10 Q8VZD2 | Q8vzd2 arabidopsis |
| 21 | 10 | 21.7 | 336 | 10 Q8S307 | Q8s307 arabidopsis |
| 22 | 10 | 21.7 | 337 | 5 Q9NJY4  | Q9njy4 drosophila  |
| 23 | 10 | 21.7 | 337 | 5 Q9NB05  | Q9nb05 drosophila  |
| 24 | 10 | 21.7 | 364 | 4 Q9BWF3  | Q9bwf3 homo sapien |
| 25 | 10 | 21.7 | 365 | 11 Q9NMA9 | Q9nma9 mus musculu |
| 26 | 10 | 21.7 | 368 | 4 Q9BTW7  | Q9btw7 homo sapien |
| 27 | 10 | 21.7 | 374 | 10 P93423 | P93423 oryza sativ |
| 28 | 10 | 21.7 | 375 | 10 P93424 | P93424 oryza sativ |
| 29 | 10 | 21.7 | 400 | 5 Q26652  | Q26652 strongyloce |
| 30 | 10 | 21.7 | 404 | 11 Q9QZW9 | Q9qzw9 mus musculu |
| 31 | 10 | 21.7 | 404 | 16 Q07801 | Q07801 mycobacteri |
| 32 | 10 | 21.7 | 421 | 5 Q93119  | Q93119 antheraea p |
| 33 | 10 | 21.7 | 425 | 3 Q12741  | Q12741 blastocladi |
| 34 | 10 | 21.7 | 431 | 5 Q9VEP2  | Q9vfp2 drosophila  |
| 35 | 10 | 21.7 | 436 | 5 Q967F8  | Q967f8 antheraea p |
| 36 | 10 | 21.7 | 446 | 4 Q9NP49  | Q9np49 homo sapien |
| 37 | 10 | 21.7 | 452 | 4 Q96MJ1  | Q96mj1 homo sapien |
| 38 | 10 | 21.7 | 462 | 5 Q9W5X7  | Q9w5x7 drosophila  |
| 39 | 10 | 21.7 | 481 | 4 Q9NW07  | Q9nw07 homo sapien |
| 40 | 10 | 21.7 | 482 | 4 Q9UKL0  | Q9ukl0 homo sapien |
| 41 | 10 | 21.7 | 484 | 11 Q923T4 | Q923t4 mus musculu |
| 42 | 10 | 21.7 | 494 | 5 Q8S224  | Q8sz24 drosophila  |
| 43 | 10 | 21.7 | 512 | 5 Q26452  | Q26452 drosophila  |
| 44 | 10 | 21.7 | 526 | 5 Q9VXV2  | Q9vxv2 drosophila  |
| 45 | 10 | 21.7 | 536 | 5 Q9W0Z3  | Q9w0z3 drosophila  |

ALIGNMENTS

RESULT 1

ID Q9TUC9 PRELIMINARY; PRT; 130 AA.  
AC Q9TUC9;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Transcription factor HOXA13 (Fragment).  
GN HOXA13.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20122168; PubMed=10656931;  
RA Mortlock D.P., Sateesh P., Innis J.W.;  
RT "Evolution of N-terminal sequences of the vertebrate HOXA13 protein.";  
EMBL; AF083096; AAD54641.1; -  
DR InterPro; IPR001859; Ribosomal\_P2.  
DR PRINTS; PR00456; RIBOSOMALP2.  
FT NON\_TER 1  
FT NON\_TER 130 130  
SQ SEQUENCE 130 AA; 10975 MW; CCA1A7B52221394 CRC64;

Query Match 21.7%; Score 10; DB 6; Length 130;  
Best Local Similarity 13.3%; Pred. No. 1.6e-121;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GXXXXXXXXXXXX 18  
|  
Db 15 GAAAAAAAAAAAAA 29

RESULT 2

ID Q9TST8 PRELIMINARY; PRT; 131 AA.  
AC Q9TST8;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Transcription factor HOXA13 (Fragment).  
 GN HOXA13  
 OS Felis silvestris catus (Cat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
 OX NCBI\_TaxID=9685;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=20122168; PubMed=10656931;  
 RA Mortlock D.P., Sateesh P., Innis J.W.;  
 RT "Evolution of N-terminal sequences of the vertebrate HOXA13 protein."  
 RL Mamm. Genome 11:151-158(2000).  
 DR EMBL; AF083095; AAD54640.1; -;  
 DR InterPro; IPR001859; Ribosomal\_P2.  
 DR PRINTS; PR00456; RIBOSOMALP2.  
 FT NON\_TER 1 131  
 SEQUENCE 131 AA; 11046 MW; 950562880E529D4F CRC64;  
 Query Match 21.7%; Score 10; DB 6; Length 131;  
 Best Local Similarity 13.3%; Pred. No. 1.6e-121;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 Qy 4 GXXXXXXXXXXXXA 18  
 Db 15 GAAAAAATAAAAAA 29  
 RESULT 3  
 ID Q9Y4M1 PRELIMINARY; PRT; 150 AA.  
 AC Q9Y4M1;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE Hypothetical 15.1 kDa protein (Fragment).  
 GN DKFZP586E1621.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;  
 RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 EMBL; AL080235; CAB45781.1; -;  
 KW Hypothetical protein.  
 FT NON\_TER 1 1  
 SEQUENCE 150 AA; 15110 MW; B0C80E46FCAB03E CRC64;  
 Query Match 21.7%; Score 10; DB 4; Length 150;  
 Best Local Similarity 13.3%; Pred. No. 1.6e-121;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 Qy 4 GXXXXXXXXXXXXA 18  
 Db 126 GTTAAAAAATAAAAAA 140  
 RESULT 4  
 ID Q9SVQ0 PRELIMINARY; PRT; 151 AA.  
 AC Q9SVQ0;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Fibroin heavy chain (Fragment).  
 GN FIB-H.  
 OS Antheraea yamamai.

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Bombycoidea; Saturniidae; Saturniini; Antheraea.  
 OX NCBI\_TaxID=7121;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Zurovec M., Yang C., Sehnel F.;  
 RT "Divergence of lepidopteran fibroin structure compatible with silk  
 thread formation."  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Fedic R.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF410906; AAL02118.1; -;  
 FT NON\_TER 1 1  
 SEQUENCE 151 AA; 13629 MW; 596775A00040475A CRC64;  
 Query Match 21.7%; Score 10; DB 5; Length 151;  
 Best Local Similarity 13.3%; Pred. No. 1.6e-121;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 Qy 4 GXXXXXXXXXXXXA 18  
 Db 39 GSSAAAAAATAAAAAA 53  
 RESULT 5  
 ID Q96N13 PRELIMINARY; PRT; 191 AA.  
 AC Q96N13;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE CDNA FLJ30829 fis, clone FEBRA2001790, highly similar to Xenopus  
 DE laevis RRM-containing protein SEB-4 mRNA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,  
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,  
 RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,  
 RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK055391; BAB70914.1; -;  
 DR InterPro; IPR001412; tRNA-synt.1.  
 DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; UNKNOWN.1.  
 RP SEQUENCE 191 AA; 19599 MW; 89EAD14E3DA941F7 CRC64;  
 Query Match 21.7%; Score 10; DB 4; Length 191;  
 Best Local Similarity 13.3%; Pred. No. 1.6e-121;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 Qy 4 GXXXXXXXXXXXXA 18  
 Db 162 GTAAAAAATAAAAAA 176  
 RESULT 6  
 ID Q9D6J0 PRELIMINARY; PRT; 193 AA.  
 AC Q9D6J0;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE Adult male hippocampus cDNA, RIKEN full-length enriched library.



|   |  |
|---|--|
| DR  | PROSITE; PS00032; ANTENNAPEDIA; UNKNOWN_1.         |
| DR  | PROSITE; PS00027; HOMEBOX_1; 1.                    |
| DR  | PROSITE; PS00071; HOMEBOX_2; 1.                    |
| KW  | DNA-binding; Homeobox; Nuclear protein.            |
| SQ  | SEQUENCE 200 AA; 22398 MW; EAC2DCA86F54E7E3 CRC64; |
| <br>Query Match 21.7%; Score 10; DB 5; Length 200;<br>Best Local Similarity 13.3%; Pred. No. 1.6e-121;<br>Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;   |  |
| QY  | 4 GXNXXXXXXXXXXA 18<br>                            |
| Db  | 29 GTSASASSAAAAAA 43                               |
| <br>RESULT 8<br>Q9GZ03 PRELIMINARY; PRT; 202 AA.<br>ID Q9GZ03 AC Q9GZ03<br>DT 01-MAR-2001 (TREMBlrel. 16, Created)<br>DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)<br>DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)<br>DE MAB-5.<br>OS Caenorhabditis briggsae.<br>OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;<br>OC Rhabditiidae; Peloderinae; Caenorhabditis.<br>OX NCBI_TaxId=6238;<br>RN [1]<br>RP SEQUENCE FROM N.A.<br>RA Grandien K., Sommer R.J.;<br>RT "Cloning and prediction of Caenorhabditis briggsae mab-5 cDNA.";<br>RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.<br>CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).<br>DR EMBL; AF277989; AAC00457.1; -.<br>DR HSSP; P02833; 9ANT.<br>DR TRANSFAC; T03749; -.<br>DR InterPro; IPR001827; Antennapedia.<br>DR InterPro; IPR000104; Antifreeze_1.<br>DR InterPro; IPR001356; Homeobox.<br>DR InterPro; IPR000047; HTH_repressor.<br>DR Pfam; PF00046; homeobox; 1.<br>DR PRINTS; PR00025; ANTENNAPEDIA.<br>DR PRINTS; PR00308; ANTFREEZEI.<br>DR PRINTS; PR00024; HOMEBOX.<br>DR PRINTS; PR00031; HTHREPRESSR.<br>DR PRODOM; PD000010; Homeobox; 1.<br>DR SMART; SM00389; HOX; 1.<br>DR PROSITE; PS00032; ANTENNAPEDIA; UNKNOWN_1.<br>DR PROSITE; PS00027; HOMEBOX_1; 1.<br>DR PROSITE; PS00071; HOMEBOX_2; 1.<br>KW DNA-binding; Homeobox; Nuclear protein.<br>SQ SEQUENCE 202 AA; 22533 MW; 9BE245E54344596D CRC64; |  |
| <br>Query Match 21.7%; Score 10; DB 5; Length 202;<br>Best Local Similarity 13.3%; Pred. No. 1.6e-121;<br>Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;   |  |
| QY  | 4 GXNXXXXXXXXXXA 18<br>                            |
| Db  | 29 GTSASSTASSTAATAA 43                             |
| <br>RESULT 9<br>Q9TUC8 PRELIMINARY; PRT; 213 AA.<br>ID Q9TUC8 AC Q9TUC8<br>DT 01-MAY-2000 (TREMBlrel. 13, Created)<br>DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)<br>DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)<br>DE Transcription factor HOXA13 (Fragment).<br>GN HOXA13.<br>OS Monodelphis domestica (Short-tailed grey opossum);<br>OC Eukaryota; Metazoa; Chordata; Graziata; Vertebrata; Euteleostomi;   |  |

OC Mammalia; Metatheria; Didelphimorphia; Didelphidae; Monodelphis.  
 OX NCBI\_TaxID=13616;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20122168; PubMed=10656931;  
 RA Morlock D.P., Satesh P., Innis J.W.;  
 RT "Evolution of N-terminal sequences of the vertebrate HOXA13 protein."  
 RL Mamm. Genome 11:151-158(2000).  
 DR EMBL; AF083097; AAD54642.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 213  
 FT NON\_TER 213  
 SQ SEQUENCE 213 AA; 19165 MW; EDEC8B40FCACADE9 CRC64;

Query Match 21.7%; Score 10; DB 6; Length 213;  
 Best Local Similarity 13.3%; Pred. No. 1.6e-121;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18  
 |  
 15 GAAAAAAAAAAAAA 29

RESULT 10  
 Q96GR0 PRELIMINARY; PRT; 218 AA.  
 AC Q96GR0;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Unknown (protein for MGC:16644).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=WUSCLE;  
 RA Strausberg R.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC009294; AAH09294.1; -.  
 DR InterPro; IPR000302; KOW\_motif.  
 DR InterPro; IPR002784; Ribosomal\_L14e.  
 DR Pfam; PF00467; KOW; 1.  
 DR Pfam; PF01929; Ribosomal\_L14e; 1.  
 SQ SEQUENCE 218 AA; 23645 MW; 3DE5CB1F97345E78 CRC64;

Query Match 21.7%; Score 10; DB 4; Length 218;  
 Best Local Similarity 13.3%; Pred. No. 1.6e-121;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18  
 |  
 148 GTAAAAAAAAAAAAA 162

RESULT 11  
 Q8WU0 PRELIMINARY; PRT; 219 AA.  
 AC Q8WU0;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Ribosomal protein L14.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=KIDNEY;  
 RA Strausberg R.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC019651; AAH19651.1; -.

DR InterPro; IPR000104; Antifreeze\_1.  
 DR InterPro; IPR000302; KOW\_motif.  
 DR InterPro; IPR002784; Ribosomal\_L14e.  
 DR Pfam; PF00467; KOW; 1.  
 DR Pfam; PF01929; Ribosomal\_L14e; 1.  
 DR PRINTS; PR00308; ANTIFREEZE1.  
 SQ SEQUENCE 219 AA; 23716 MW; 81541BAF7177AB91 CRC64;

Query Match 21.7%; Score 10; DB 4; Length 219;  
 Best Local Similarity 13.3%; Pred. No. 1.6e-121;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18  
 |  
 148 GTAAAAAAAAAAAAA 162

RESULT 12  
 Q9BSB8 PRELIMINARY; PRT; 222 AA.  
 AC Q9BSB8;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Similar to ribosomal protein L14.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=OVARY;  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC005134; AAH05134.1; -.  
 DR InterPro; IPR000104; Antifreeze\_1.  
 DR InterPro; IPR000302; KOW\_motif.  
 DR Pfam; PF00467; KOW; 1.  
 DR Pfam; PF01929; Ribosomal\_L14e; 1.  
 DR PRINTS; PR00308; ANTIFREEZE1.  
 KW Ribosomal protein.  
 SQ SEQUENCE 222 AA; 23945 MW; ED7BEA42BEE5F3F CRC64;

Query Match 21.7%; Score 10; DB 4; Length 222;  
 Best Local Similarity 13.3%; Pred. No. 1.6e-121;  
 Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GXXXXXXXA 18  
 |  
 148 GTAAAAAAAAAAAAA 162

RESULT 13  
 Q9ERQ7 PRELIMINARY; PRT; 224 AA.  
 AC Q9ERQ7;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Homeodomain-containing transcription factor Nkx6.1 (Fragment).  
 GN NKX6-1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV;  
 RX MEDLINE=20507862; PubMed=10938085;  
 RA Watada H., Mirmira R.G., Leung J., German M.S.;  
 RT "Transcriptional and translational regulation of beta-cell differentiation factor Nkx6.1";

RL J. Biol. Chem. 275:34224-34230(2000).  
DR EMBL; AF291666; AAC30415.1; -.  
DR TRANSFAC; T04269; -.  
DR MGD; MGI:1206039; Nkx6-1.  
DR InterPro; IPR00104; Antifreeze\_1.  
DR PRINTS; PR00308; ANTIFREEZE1.  
FT NON\_TER 224 224  
SQ SEQUENCE 224 AA; 648418317532D3A9 CRC64;

Query Match 21.7%; Score 10; DB 11; Length 224;  
Best Local Similarity 13.3%; Pred. No. 1.6e-121;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GXXXXXXXXXXXX 18  
Db 122 GSSSSSSSSASATS 136

## RESULT 14

Q9BX46 PRELIMINARY; PRT; 236 AA.  
Q9BX46;  
DT 01-JUN-2001 (TRENBLrel. 17, Created)  
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE DJ259A10.1 (ssDNA binding protein (SEB4D)).  
GN DJ259A10.1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Babbage A.;  
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AL136305; CAC36889.1; -.  
DR HSSP; P09651; 1UP1.  
DR InterPro; IPR00104; Antifreeze\_1.  
DR InterPro; IPR01778; POA\_allergenC.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR InterPro; IPR001412; tRNA-synt\_1.  
DR Pfam; PF00076; rim; 1.  
DR PRINTS; PR00308; ANTIFREEZE1.  
DR PRINTS; PR00833; POAALLERGEN.  
DR SMART; SM00360; RM; 1.  
DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; UNKNOWN\_1.  
DR PROSITE; PS50102; RRM; 1.  
DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
SQ SEQUENCE 236 AA; 24776 MW; 1CFB5AEBD4E3AA24 CRC64;

Query Match 21.7%; Score 10; DB 4; Length 236;  
Best Local Similarity 13.3%; Pred. No. 1.6e-121;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GXXXXXXXXXXXX 18  
Db 207 GTAAAAAAXXXXXX 221

## RESULT 15

Q91XV7 PRELIMINARY; PRT; 285 AA.  
AC Q91XV7;  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE Brain specific binding protein.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=WISTERST; TISSUE=HIPPOCAMPUS;  
RX MEDLINE=21402891; PubMed=11399754;  
RA Hama T., Maruyama M., Katoh-Semba R., Takizawa M., Iwashima M.,  
RA Nara K.;  
RT "Identification and Molecular Cloning of a Novel Brain-specific  
RT Receptor Protein That Binds to Brain Injury-derived Neurotrophic  
RT Peptide. Possible Role for Neuronal Survival.";  
RL J. Biol. Chem. 276:31929-31935(2001).  
DR EMBL; AB028891; BAB63459.1; -.  
SQ SEQUENCE 285 AA; 29004 MW; E1815ED358A07AF7 CRC64;

Query Match 21.7%; Score 10; DB 11; Length 285;  
Best Local Similarity 13.3%; Pred. No. 1.6e-121;  
Matches 2; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 GXXXXXXXXXXXX 18  
Db 260 GTTAAAAAAXXXXXX 274

Search completed: January 8, 2003, 11:00:51  
Job time : 25.6034 secs



OM protein - protein search, using sw model  
 Run on: January 8, 2003, 10:54:49 ; Search time 52.4483 Seconds  
 (without alignments)  
 99.084 Million cell updates/sec

Title: US-09-003-869-5  
 Perfect score: 42  
 Sequence: 1 XX 39

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5  
 Searched: 908470 seqs, 133250620 residues  
 al number of hits satisfying chosen parameters: 908470  
 Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000  
 Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 4     | 9.5         | 2      | 21 | AA49410     |
| 2          | 4     | 9.5         | 2      | 22 | AA49410     |
| 3          | 4     | 9.5         | 3      | 2  | AA49410     |
| 4          | 4     | 9.5         | 3      | 4  | AA49410     |
| 5          | 4     | 9.5         | 3      | 6  | AA49410     |
| 6          | 4     | 9.5         | 3      | 10 | AA49410     |
| 7          | 4     | 9.5         | 3      | 10 | AA49410     |
| 8          | 4     | 9.5         | 3      | 10 | AA49410     |
| 9          | 4     | 9.5         | 3      | 10 | AA49410     |
| 10         | 4     | 9.5         | 3      | 14 | AA49410     |

# SUMMARIES

|    |   |     |   |    |         |
|----|---|-----|---|----|---------|
| 11 | 4 | 9.5 | 3 | 14 | AA49410 |
| 12 | 4 | 9.5 | 3 | 14 | AA49410 |
| 13 | 4 | 9.5 | 3 | 15 | AA49410 |
| 14 | 4 | 9.5 | 3 | 15 | AA49410 |
| 15 | 4 | 9.5 | 3 | 16 | AA49410 |
| 16 | 4 | 9.5 | 3 | 16 | AA49410 |
| 17 | 4 | 9.5 | 3 | 16 | AA49410 |
| 18 | 4 | 9.5 | 3 | 16 | AA49410 |
| 19 | 4 | 9.5 | 3 | 16 | AA49410 |
| 20 | 4 | 9.5 | 3 | 16 | AA49410 |
| 21 | 4 | 9.5 | 3 | 16 | AA49410 |
| 22 | 4 | 9.5 | 3 | 16 | AA49410 |
| 23 | 4 | 9.5 | 3 | 16 | AA49410 |
| 24 | 4 | 9.5 | 3 | 16 | AA49410 |
| 25 | 4 | 9.5 | 3 | 16 | AA49410 |
| 26 | 4 | 9.5 | 3 | 17 | AA49410 |
| 27 | 4 | 9.5 | 3 | 18 | AA49410 |
| 28 | 4 | 9.5 | 3 | 19 | AA49410 |
| 29 | 4 | 9.5 | 3 | 19 | AA49410 |
| 30 | 4 | 9.5 | 3 | 19 | AA49410 |
| 31 | 4 | 9.5 | 3 | 19 | AA49410 |
| 32 | 4 | 9.5 | 3 | 19 | AA49410 |
| 33 | 4 | 9.5 | 3 | 19 | AA49410 |
| 34 | 4 | 9.5 | 3 | 20 | AA49410 |
| 35 | 4 | 9.5 | 3 | 20 | AA49410 |
| 36 | 4 | 9.5 | 3 | 20 | AA49410 |
| 37 | 4 | 9.5 | 3 | 20 | AA49410 |
| 38 | 4 | 9.5 | 3 | 20 | AA49410 |
| 39 | 4 | 9.5 | 3 | 20 | AA49410 |
| 40 | 4 | 9.5 | 3 | 20 | AA49410 |
| 41 | 4 | 9.5 | 3 | 20 | AA49410 |
| 42 | 4 | 9.5 | 3 | 20 | AA49410 |
| 43 | 4 | 9.5 | 3 | 20 | AA49410 |
| 44 | 4 | 9.5 | 3 | 20 | AA49410 |
| 45 | 4 | 9.5 | 3 | 20 | AA49410 |

## ALIGNMENTS

RESULT 1  
 AA49410  
 ID AA49410 standard; peptide; 2 AA.  
 XX AA49410;  
 AC AA49410;  
 DT 13-MAR-2000 (first entry)  
 XX Novel autism peptide as diagnostic marker.  
 DE Diagnostic marker; human disorder; opiate; autism spectral disorder;  
 KW autism pervasive developmental disorder; Asperger's syndrome;  
 KW attention deficient disorder; attention hyperactivity disorder;  
 KW multiple sclerosis; Parkinson's disease; Alzheimer's dementia.  
 XX Synthetic.  
 OS Homo sapiens.  
 XX EP969015-A2.  
 XX 05-JAN-2000.  
 XX 15-JUN-1999; 99EP-0304636.  
 XX 15-JUN-1998; 98US-0089237.  
 PR 15-JUN-1998; 98US-0089238.  
 PR 24-MAY-1999; 99US-0317702.  
 XX (ORTH-) ORTHO-CLINICAL DIAGNOSTICS INC.  
 PA Shanahan MR, Venturini AJ, Daiss JL, Friedman AE;  
 PI WPI; 2000-074781/07.

Osteogenic protein  
 Chromophore-contg.  
 Phytase derived pe  
 Calcium ion channe  
 Microbial emetic t  
 Apolipoprotein A-I  
 Apolipoprotein A-I  
 Apolipoprotein A-I  
 Apolipoprotein A-I  
 Apolipoprotein A-I  
 DP-178 homologue 1  
 HPF3 peptide deriv  
 Beta-amyloid modul  
 Human IL-16 5' end  
 IDA-peptide capabl  
 Fusion immunoglobu  
 Anti-inflammatory  
 Anti-inflammatory  
 Anti-inflammatory  
 Loop region used i  
 Linker for dual av  
 Elastin peptide fr  
 Peptide linker use  
 Peptide comprising  
 Peptide comprising  
 Apolipoprotein A-I  
 Apolipoprotein A-I  
 Apolipoprotein A-I  
 Apolipoprotein A-I  
 Apolipoprotein A-I

XX PT Diagnosing human disorders e.g. autism spectral disorders, multiple  
 PT sclerosis, Parkinson's disease and Alzheimer's dementia -  
 XX Claim 9; Page 8; 44pp; English.  
 XX The invention provides diagnostic markers for a human disorder,  
 CC comprising either opiate-like peptides or opiate-derived peptides. The  
 CC novel peptides are used as diagnostic marker, in ex-vivo methods of  
 CC diagnosing human disorders e.g. autism spectral disorders including  
 CC autism pervasive developmental disorder, Aspergers syndrome, attention  
 CC deficient disorder and attention hyperactivity disorder, and multiple  
 CC sclerosis, Parkinson's disease and Alzheimer's dementia. Sequences  
 CC AX49371-413 represent peptide diagnostic markers for the human  
 CC disorders specified above.  
 XX Sequence 2 AA;  
 SQ Query Match 9.5%; Score 4; DB 21; Length 2;  
 Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 18 A 18  
 Db 1 A 1  
 RESULT 2  
 AAG99822 standard; Peptide; 2 AA.  
 ID AAG99822  
 XX AAG99822;  
 XX 27-SEP-2001 (first entry)  
 DT ERA binding domain polypeptide SEQ ID NO 264.  
 DE ERA binding domain; Escherichia coli; GTPase; antimicrobial;  
 KW antibacterial; antibiotic; pathogenesis; infection; vaccine;  
 KW peptide therapy.  
 XX Homo sapiens.  
 OS WO200153458-A2.  
 PN 26-JUL-2001.  
 PD 17-JAN-2001; 2001WO-US01786.  
 XX 18-JAN-2000; 2000US-0176870.  
 XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX Lupas AN, Pearce KH;  
 PI WPI; 2001-476108/51.  
 DR New ERA binding domain polypeptides and polynucleotides encoding them,  
 XX useful as research reagents and materials for discovery of treatments  
 XX and diagnostics for diseases, or for genetic immunisation -  
 XX Claim 1; Page 37; 279pp; English.  
 XX The present invention relates to ERA binding domain polypeptides  
 CC (AAG99599-AAG99989 and AAM00010-AAM00021). The era gene in Escherichia  
 CC coli codes for an essential GTPase protein able to autophosphorylate at  
 CC serine and/or threonine residues. The protein has potential antimicrobial  
 CC and antibacterial activity and is useful in screening for antagonists,  
 CC agonists and for compounds with antibiotic activity. The proteins are  
 CC also useful in determining their role in pathogenesis of infection,  
 CC dysfunction and disease and could be used as part of a vaccine and/or  
 CC peptide therapy.

XX SQ Sequence 2 AA;  
 Query Match 9.5%; Score 4; DB 22; Length 2;  
 Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 18 A 18  
 Db 1 A 1  
 RESULT 3  
 AAP10296  
 ID AAP10296 standard; Protein; 3 AA.  
 XX AAP10296;  
 AC 15-DEC-1992 (first entry)  
 DT (Ala)3-Daunomycin.  
 DE Antibiotic; cancer; anthracycline.  
 XX Synthetic.  
 XX US4256632-A.  
 PN 17-MAR-1981.  
 PD 20-AUG-1979; 79US-0067664.  
 PF 24-AUG-1978; 78IL-0055431.  
 PR (YEDA ) YEDA RES & DEV CO LTD.  
 PA Levin Y, Sela BA;  
 PI WPI; 1981-25242D/14 (25242D).  
 XX Aminoacid and peptide derivs. of daunomycin - with antitumour  
 XX activity and reduced toxicity  
 XX Disclosure; Page 3; 3pp; English.  
 XX Novel daunomycin derivs. (Ala)n-Daunomycin (n= 1-3) are obtained  
 CC by linking (Ala)n to the amino gp. of the drug mol.  
 CC Due to the reduced toxicity of the deriv. it is possible to  
 CC administer dosages of ca. 5 to 40 times larger than daunomycin itself.  
 XX Sequence 3 AA;  
 SQ Query Match 9.5%; Score 4; DB 2; Length 3;  
 Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 18 A 18  
 Db 1 A 1  
 RESULT 4  
 AAP30601  
 ID AAP30601 standard; Protein; 3 AA.  
 XX AAP30601;  
 AC 31-MAY-1992 (first entry)  
 DT Sequence of beta-lactamase encoded on pULB1523;  
 DE Emphysema therapy; proteolytic enzyme; lung disease.  
 XX

OS Homo sapiens.  
 XX BE895961-A.  
 XX 16-JUN-1983.  
 XX 06-JAN-1984; 84BE-0870005.  
 XX 21-FEB-1983; 83BE-0895961.  
 PR 21-FEB-1983; 83BE-0210157.  
 XX (REGI-) REGION WALLONNE.  
 PA (EJEC-) EJECUTIVO REG VALON.  
 XX WPI: 1983-700089/27.  
 DR N-PSDB; AAN30207.  
 XX Bacterial clone producing alpha-1-antitrypsin - transformed with  
 PT vector contg. antitrypsin DNA  
 XX Disclosure: Fig 2; 23pp; French.  
 CC The inventors claim double-stranded cDNA encoding AT and E.coli  
 CC clones contg. AT-DNA. The cDNA is derived from mRNA extracted from  
 CC human liver. Recombinant alpha-1-AT can be used for the treatment of  
 CC pollutant-induced lung damage (esp. emphysema), particularly when  
 CC caused by excessive release of proteolytic enzymes. Large amts. of  
 CC AT can be made by cultivating the transformed bacteria.  
 XX Sequence 3 AA;  
 SQ

Query Match 9.5%; Score 4; DB 4; Length 3;  
 Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 18 A 18  
 DB 2 A 2

RESULT 5  
 AAP50012  
 ID AAP50012 standard; protein; 3 AA.  
 XX  
 AC AAP50012;  
 XX  
 DT 16-AUG-2002 (updated)  
 03-SEP-1991 (first entry)  
 Sequence linking insulin B chain B(1-29) to A chain a(1-21) to form  
 a biosynthetic insulin precursor.  
 DE Insulin precursor; diabetes; proteolysis resistant.  
 XX Synthetic.  
 XX  
 OS  
 FH Key Location/Qualifiers  
 FT Misc-difference 1 /note= "attached to B(1-29)"  
 FT Misc-difference 3 /note= "attached to A(1-21)"  
 FT  
 XX EP163529-A.  
 XX  
 XX 04-DEC-1985.  
 XX  
 XX 29-MAY-1985; 85EP-0116297.  
 XX  
 XX 08-FEB-1985; 85DK-0000582.  
 PR 30-MAY-1984; 84DK-0002665.  
 PR 29-MAY-1985; 85DK-0002385.  
 PR 15-NOV-1990; 90EP-0121887.  
 XX

(NOVO ) NOVO INDUSTRI A/S.  
 PA Markussen J, Fill N, Ammerer G, Hansen M, Thim L, Norris K;  
 PI Voigt HO;  
 XX WPI: 1985-304970/49.  
 XX DNA-sequence encoding insulin precursor - having correctly  
 PT positioned di:sulphide bridges and resistant to proteolytic degradation  
 XX Claim 4; Page 32; 47pp; English.  
 XX The biosynthetic insulin precursors are generated largely with  
 CC correctly positioned disulphide bridges between the A- and B-  
 CC moieties, and are more resistant to proteolytic degradation than  
 CC previously.  
 CC (Updated on 16-AUG-2002 to add missing OS field.)  
 XX Sequence 3 AA;  
 SQ

Query Match 9.5%; Score 4; DB 6; Length 3;  
 Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 18 A 18  
 DB 1 A 1

RESULT 6  
 AAP93348  
 ID AAP93348 standard; protein; 3 AA.  
 XX  
 AC AAP93348;  
 XX  
 DT 24-JUN-1990 (first entry)  
 XX  
 DE Portion of myc variant as encoded by pMYC-A.  
 XX C-myc; variant myc protein; pMYC-A.  
 XX JP01039999-A.  
 PN 10-FEB-1989.  
 XX  
 PD 06-AUG-1987; 87JP-0197197.  
 XX  
 PF 06-AUG-1987; 87JP-0197197.  
 XX  
 PR 06-AUG-1987; 87JP-0197197.  
 XX  
 XX (MITK ) MITSUI TOATSU CHEM INC.  
 PA WPI: 1989-089714/12.  
 DR N-PSDB; AAN91367.  
 XX  
 XX Myc protein for antibody prepn. -  
 PT is stabilised by converting aminoacid(s) of myc protein to other  
 PT aminoacid to produce variant type myc protein  
 XX Fig 3.4.; page 17; 18pp; Japanese.  
 PS  
 XX It is a variant of C-myc protein. In the patent, myc protein in the host  
 CC cell is stabilised by converting at least one amino acid to another amino  
 CC acid, while not affecting the properties of the myc protein. The residue  
 CC changed is the second one.  
 XX  
 XX Sequence 3 AA;  
 SQ

Query Match 9.5%; Score 4; DB 10; Length 3;  
 Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 18 A 18  
 DB 1

Db 2 A 2

RESULT 7  
AAP90667  
ID AAP90667 standard; protein; 3 AA.

XX AC AAP90667;  
XX DT 26-MAY-1990 (first entry)  
XX DE New antihypertensive peptide.  
XX KW Antihypertensive peptide; angiotensin-converting enzyme;  
XX ACE.

PN JP01083096-A.

PD 28-MAR-1989.

25-SEP-1987; 87JP-0241646.

25-SEP-1987; 87JP-0241646.

(AJIN ) AJINOMOTO KK.

WPI; 1989-136272/18.

PT New antihypertensive peptide(s) -  
used to inhibit angiotensin-converting enzyme

XX PS ; 20pp; Japanese.

XX CC The peptide and its salts inhibit angiotensin-converting  
enzyme (ACE) and are useful as antihypertensives. They may be  
administered orally, parenterally or rectally in the form of tablets,  
capsules, granules, powder, syrup, suspension, suppositories,  
ointment, cream, gel, plaster, inhalation compsn. or  
injection at a dose of 0.001-1000, pref. 0.01-10, mg, 1-3 times per day.

SQ Sequence 3 AA;

Query Match 9.5%; Score 4; DB 10; Length 3;  
Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
2 A 2

RESULT 8  
AAP90668  
ID AAP90668 standard; protein; 3 AA.

XX AC AAP90668;  
XX DT 26-MAY-1990 (first entry)  
XX DE New antihypertensive peptide.  
XX KW Antihypertensive peptide; angiotensin-converting enzyme;  
XX ACE.

PN JP01083096-A.

PD 28-MAR-1989.

25-SEP-1987; 87JP-0241646.

25-SEP-1987; 87JP-0241646.

(AJIN ) AJINOMOTO KK.

XX WPI; 1989-136272/18.

XX New antihypertensive peptide(s) -  
used to inhibit angiotensin-converting enzyme

XX PS ; 20pp; Japanese.

XX CC The peptide and its salts inhibit angiotensin-converting  
enzyme (ACE) and are useful as antihypertensives. They may be  
administered orally, parenterally or rectally in the form of tablets,  
capsules, granules, powder, syrup, suspension, suppositories,  
ointment, cream, gel, plaster, inhalation compsn. or  
injection at a dose of 0.001-1000, pref. 0.01-10, mg, 1-3 times per day.

SQ Sequence 3 AA;

Query Match 9.5%; Score 4; DB 10; Length 3;  
Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
3 A 3

RESULT 9

AAP97811

ID AAP97811 standard; Protein; 3 AA.

XX AC AAP97811;

XX DT 29-JUL-1992 (first entry)

XX DE Sequence of fragment 24, the tryptic fragment of recombinant  
penicillin acyltransferase (PAT) polypeptide 2.

XX KW Penicillin biosynthesis; enzyme; antibiotic.

XX OS Penicillium chrysogenum.

XX PN EP336446-A.

XX PD 11-OCT-1989.

XX PF 07-APR-1989; 89EP-0106214.

XX PR 08-SEP-1988; 88AT-0002201.

XX PR 08-APR-1988; 88AT-0000922.

XX PR 13-JUL-1988; 88AT-0001806.

XX PA (BIOC ) BIOCHEMIE GMBH.

XX PI Knauseder F, Leitner E, Palma N, Weber G;

XX WPI; 1989-294357/41.

XX Recombinant penicillin acyl-transferase - and DNA coding for it

XX PS Claim 9; Page 48; 52pp; English.

XX CC The inventors claim recombinant penicillin acyltransferase (PAT) and  
DNA coding for PAT. PAT catalyses the last step in the biosynthesis  
of penicillin G and penicillin V. More specifically, the coding  
strand of the DNA has the nucleotide sequence shown below. This  
CC includes three introns and codes for a PAT protein with mol. wt. ca.  
40kD. Plasmid vectors pBC2001 and pBC2002 are specifically claimed.

XX SQ Sequence 3 AA;

Query Match 9.5%; Score 4; DB 10; Length 3;  
Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 18 A 18  
Db 1 A 1

## RESULT 10

AAR42569  
ID AAR42569 standard; peptide; 3 AA.

XX AC AAR42569;  
XX DT 22-JUN-1994 (first entry)

XX DE Peptide corresponding to pseudo-substrate region of zeta-PKC.  
XX KW Zeta-protein kinase C inhibitor; zeta-PKC; pseudosubstrate; tumour;  
XX KW hyperproliferative disorders; psoriasis; viral infection; HIV.

XX XX Synthetic.

XX Key Location/Qualifiers  
XX Modified-site 1 /note= "can be N-acetylated"  
XX WO9320101-A.  
XX 14-OCT-1993.  
XX 02-APR-1993; 93WO-EP00816.  
XX 06-APR-1992; 92EP-0500034.  
XX (GLAX ) GLAXO SA.  
XX DIAZ-MECO CONDE MT, MOSCAT GUILLEN J;  
XX WPI: 1993-336831/42.

XX Peptide(s) corresp. to the pseudo-substrate region of zeta-PKC -  
XX used for treatment of tumours, hyper-proliferative disorders and  
XX viral infections  
XX Claims 4 + 5; Page 43; 57pp; English.  
XX The main claim refers to new peptides of formula X-Ala-Arg-Arg-J in  
XX which X is H or one or more amino acids and J is OH or one or more  
XX amino acids, the peptides containing a total of 3 to 15 amino acids.  
XX The present peptide is a specifically claimed example of these new  
XX peptides  
XX The peptides are specific inhibitors of protein kinase C isotype  
XX zeta, i.e. any subspecies of PKC which contains the specific  
XX autoinhibitory pseudosubstrate domain RRGARRWRK (Acc. No. AAR42573).  
XX This domain has been found to be perfectly conserved in zeta-PKC  
XX variants isolated from a number of different sources, including rat  
XX brain. The peptides are usefully therapeutically for treating  
XX conditions where the underlying aetiology is associated with  
XX zeta-PKC, including tumours, hyperproliferative disorders (e.g.  
XX psoriasis) and viral infections (e.g. HIV).

XX Sequence 3 AA;

Query Match 9.5%; Score 4; DB 14; Length 3;  
Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 18 A 18  
Db 1 A 1

## RESULT 11

AAR53397

ID AAR53397 standard; Protein; 3 AA.

XX AC AAR53397;  
XX DT 01-JUL-2002 (updated)

XX DT 06-JUN-1994 (first entry)  
XX DE Osteogenic protein fragment.

XX KW Osteogenic protein; bone; cartilage; matrix; osteoarthritis;  
XX KW repair; vascularisation; mineralisation; differentiation.  
XX OS Mammalia.

XX PN US5266683-A.

XX PD 30-NOV-1993.

XX PF 08-APR-1988; 88US-0179406.

XX PR 08-APR-1988; 88US-0179406.

XX PR 15-AUG-1988; 88US-0232630.

XX PR 23-FEB-1989; 89US-0315342.

XX PR 17-OCT-1989; 89US-0422613.

XX PR 17-OCT-1989; 89US-0422699.

XX PR 22-FEB-1990; 90US-0483913.

XX PR 20-AUG-1990; 90US-0569920.

XX PR 07-SEP-1990; 90US-0579865.

XX PR 18-OCT-1990; 90US-0599543.

XX PR 18-OCT-1990; 90US-0600024.

XX PR 04-DEC-1990; 90US-0621849.

XX PR 22-FEB-1991; 91US-0660162.

XX PR 20-DEC-1991; 91US-0810560.

XX PR 28-JAN-1992; 92US-0827052.

XX PR 21-FEB-1992; 92US-0841646.

XX (STYC ) STRYKER CORP.

XX Kuberampath T, Oppermann H, Ozkaynak E, Pang RHL;

XX Rueger DC;

XX WPI: 1993-395405/49.

XX PT New pure mammalian osteogenic proteins - induce cartilage and  
XX endochondral bone formation when in association with a matrix

XX Claim 1; Column 153; 128pp; English.

XX Osteogenic proteins when in association with a matrix can induce  
XX at the locus of an implant the full development cascade of  
XX endochondral bone formation including vascularisation,  
XX mineralisation and bone marrow differentiation. They can also be  
XX used to repair both bone and cartilage in the treatment of  
XX osteoarthritis.  
XX (Updated on 01-JUL-2002 to add missing PA field.)

XX Sequence 3 AA;

Query Match 9.5%; Score 4; DB 14; Length 3;  
Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 18 A 18  
Db 1 A 1

## RESULT 12

AAR34623

ID AAR34623 standard; peptide; 3 AA.

XX AC AAR34623;

```

XX DT 12-NOV-1992 (first entry)
XX DE Chromophore-contg. cpd. for determining ICE activity (3).
XX KW Interleukin-beta; IL-1b; inflammation; immune; disease;
XX KW diagnosis; 7-amino-4-methylcoumarin amide.
XX OS Synthetic.
XX PH Key Location/Qualifiers
FT Modified-site 1
FT Modified-site 3 /note= "N-acetyl-tyrosinyl"
FT Modified-site 3 /note= "alaninyl 7-amino-4-methylcoumarin amide"
XX EP528487-A.
XX 24-FEB-1993.
XX 08-AUG-1992; 92EP-0202450.
PR 16-AUG-1991; 91US-0746455.
PR 17-DEC-1991; 91US-0808994.
XX (MERI ) MERCK & CO INC.
XX Chapman KT, Hagmann WK, Maccoss M, Mumford RA, Thornberry NA;
XX Weidner JR;
XX WPI; 1993-060350/08.
XX New chromophore-contg. cpds. - for determining interleukin-beta
XX PT convertase activity in diagnosis of inflammatory or immune-based
XX PT disorders
XX PS Claim 9; Page 42; 42pp; English.
XX The cpd. is used for determining interleukin-beta convertase (ICE)
XX CC activity. ICE has been implicated in inflammatory and immune-based
XX CC diseases including diseases of the lungs and airways, CNS, eyes,
XX CC ears, joints, bones and connective tissues, cardiovascular system
XX CC (including the pericardium), GI and urogenital systems and skin and
XX CC mucosal membranes.
XX SQ Sequence 3 AA;

Query Match 9.5%; Score 4; DB 14; Length 3;
Best Local Similarity 100.0%; Pred. No. 7.1e+05;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
DB 3 A 3

RESULT 13
AAR46817
ID AAR46817 standard; Protein: 3 AA.
XX AC AAR46817;
XX DT 19-AUG-1994 (first entry)
XX DE Phytase derived peptide (15 phy).
XX KW pH 2.5; acid phosphatase; Trichoderma; host; Aspergillus; phytic acid;
XX KW phytate degrading enzyme; PDE; removal; inositol hexaphosphoric acid;
XX KW plant; feed composition; filtration.
XX OS Synthetic.
XX PN WO9403612-A.

Query Match 9.5%; Score 4; DB 15; Length 3;
Best Local Similarity 100.0%; Pred. No. 7.1e+05;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
DB 3 A 3

RESULT 14
AAR46824
ID AAR46824 standard; Protein: 3 AA.
XX AC AAR46824;
XX DT 19-AUG-1994 (first entry)
XX DE Phytase derived peptide C-terminal (C phy).
XX KW pH 2.5; acid phosphatase; Trichoderma; host; Aspergillus; phytic acid;
XX KW phytate degrading enzyme; PDE; removal; inositol hexaphosphoric acid;
XX KW plant; feed composition; filtration.
XX OS Synthetic.
XX PN WO9403612-A.
XX PD 17-FEB-1994.
XX PF 30-JUL-1993; 93WO-FI00310.
XX PR 31-JUL-1992; 92US-0923724.
XX PA (ALKO-) ALKO LTD.

```

```

XX PD 17-FEB-1994.
XX PF 30-JUL-1993; 93WO-FI00310.
XX PR 31-JUL-1992; 92US-0923724.
XX PA (ALKO-) ALKO LTD.
XX PI Cantrell M, Fagerstrom RB, Miettinen-Oinonen ASK;
XX PI Nevalainen HK, Palohelmo MT, Piddington C, Rambosek JA;
XX PI Torkkeli TK, Turunen MK;
XX DR WPI; 1994-065700/08.
XX CC Camps. contg. phytate degrading enzymes - obtd. by expression
XX PT of their genes in Trichoderma, used partic. for producing animal
XX PT feed compsns.
XX PS Example 4; Page 44; 142pp; English.
XX CC The sequences given in AAR46793-824 are peptides derived from the
XX CC phytase protein. The phytase protein may be used in the composition
XX CC of the invention. The DNA encoding the phytase protein may be
XX CC introduced into a Trichoderma host which then expresses it and the
XX CC protein is collected from the culture medium. By using Trichoderma as
XX CC a host for Aspergillus phytate degrading enzymes such as this, a
XX CC totally different enzyme composition compared to that secreted from
XX CC Aspergillus results. The enzyme composition can be used for removal
XX CC of phytic acid or inositol hexaphosphoric acid from raw material,
XX CC particularly plant material. The composition is used in feed
XX CC compositions for animals. By using Trichoderma as a source of a
XX CC composition containing phytate degrading enzymes some difficult
XX CC downstream processing problems, eg. filtration, that occur with
XX CC similar Aspergillus compositions are avoided and yields are improved.
XX SQ Sequence 3 AA;

Query Match 9.5%; Score 4; DB 15; Length 3;
Best Local Similarity 100.0%; Pred. No. 7.1e+05;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
DB 3 A 3

RESULT 14
AAR46824
ID AAR46824 standard; Protein: 3 AA.
XX AC AAR46824;
XX DT 19-AUG-1994 (first entry)
XX DE Phytase derived peptide C-terminal (C phy).
XX KW pH 2.5; acid phosphatase; Trichoderma; host; Aspergillus; phytic acid;
XX KW phytate degrading enzyme; PDE; removal; inositol hexaphosphoric acid;
XX KW plant; feed composition; filtration.
XX OS Synthetic.
XX PN WO9403612-A.
XX PD 17-FEB-1994.
XX PF 30-JUL-1993; 93WO-FI00310.
XX PR 31-JUL-1992; 92US-0923724.
XX PA (ALKO-) ALKO LTD.

```

PI Cantrell M, Fagerstroem RB, Miettinen-Oinonen ASK;  
PI Navalainen HK, Palohelmo MT, Piddington C, Rambousek JA;  
PI Torkkeli TK, Turunen MK;  
XX WPI: 1994-065700/08.

XX Compns. contg. phytate degrading enzymes - obtd. by expression  
XX of their genes in Trichoderma, used partic. for producing animal  
XX feed compns.

XX Example 4; Page 45; 142pp; English.

XX The sequences given in AAR46793-824 are peptides derived from the  
XX phytase protein. The phytase protein may be used in the composition  
XX of the invention. The DNA encoding the phytase protein may be  
XX introduced into a Trichoderma host which then expresses it and the  
XX protein is collected from the culture medium. By using Trichoderma as  
XX a host for Aspergillus phytate degrading enzymes such as this, a  
XX totally different enzyme composition compared to that secreted from  
XX Aspergillus results. The enzyme composition can be used for removal  
XX of phytic acid or inositol hexaphosphoric acid from raw material,  
XX particularly plant material. The composition is used in feed  
XX compositions for animals. By using Trichoderma as a source of a  
XX composition containing phytate degrading enzymes some difficult  
XX downstream processing problems, eg. filtration, that occur with  
XX similar Aspergillus compositions are avoided and yields are improved.

XX Sequence 3 AA;

Query Match 9.5%; Score 4; DB 15; Length 3;  
Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18  
|  
Db 3 A 3

#### RESULT 15

AAR85069  
ID AAR85069 standard; peptide; 3 AA.

XX AC AAR85069;

XX DT 13-JUN-1996 (first entry)

XX DE Calcium ion channel blocking peptide fragment #2.

Calcium ion channel; blocking peptide; hypotensive agent; blood pressure;  
therapy.

XX OS Synthetic.

XX PN JP07278185-A.

XX PD 24-OCT-1995.

XX PF 08-APR-1994; 94JP-0070579.

XX PR 08-APR-1994; 94JP-0070579.

XX PA (FARH ) HOECHST JAPAN KK.

XX DR WPI: 1995-400979/51.

XX Peptide having calcium channel blocking activity - useful in  
XX hypotensive agent

XX PS Claim 2; Page 2; 6pp; Japanese.

XX AAR85068 and AAR85069 represent calcium ion channel blocking peptides.  
XX This sequence corresponds to residues 10-13 of the peptide represented  
XX by AAR85068. The peptides suppress blood pressure by inhibiting the

CC calcium influx into cells. These peptides can be used in a hypotensive  
CC agent.

SQ Sequence 3 AA;

Query Match 9.5%; Score 4; DB 16; Length 3;  
Best Local Similarity 100.0%; Pred. No. 7.1e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18  
|  
Db 2 A 2

Search completed: January 8, 2003, 10:58:24  
Job time : 53.4483 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2003, 10:54:49 ; Search time 11.0948 Seconds  
(without alignments)  
103.426 Million cell updates/sec

Title: US-09-003-869-5  
Perfect score: 42  
Sequence: 1 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 39

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues  
Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA: \*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/iaa/FCUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| SUMMARIES  |       |               |        |    |                   |                    |  |  |  |  |  |
|------------|-------|---------------|--------|----|-------------------|--------------------|--|--|--|--|--|
| Result No. | Score | Query Match % | Length | ID | Description       |                    |  |  |  |  |  |
| 1          | 4     | 9.5           | 2      | 1  | US-07-914-280-1   | Sequence 1, Appli  |  |  |  |  |  |
| 2          | 4     | 9.5           | 2      | 1  | US-07-791-213D-23 | Sequence 23, Appli |  |  |  |  |  |
| 3          | 4     | 9.5           | 2      | 1  | US-07-840-077A-5  | Sequence 5, Appli  |  |  |  |  |  |
| 4          | 4     | 9.5           | 2      | 1  | US-08-122-510-10  | Sequence 10, Appli |  |  |  |  |  |
| 5          | 4     | 9.5           | 2      | 1  | US-08-358-160-169 | Sequence 169, App  |  |  |  |  |  |
| 6          | 4     | 9.5           | 2      | 1  | US-08-358-160-170 | Sequence 170, App  |  |  |  |  |  |
| 7          | 4     | 9.5           | 2      | 1  | US-08-454-950-5   | Sequence 5, Appli  |  |  |  |  |  |
| 8          | 4     | 9.5           | 2      | 1  | US-08-454-949-5   | Sequence 5, Appli  |  |  |  |  |  |
| 9          | 4     | 9.5           | 2      | 1  | US-08-153-799-20  | Sequence 20, Appli |  |  |  |  |  |
| 10         | 4     | 9.5           | 2      | 1  | US-08-293-150A-23 | Sequence 23, Appli |  |  |  |  |  |
| 11         | 4     | 9.5           | 2      | 2  | US-08-727-688-22  | Sequence 22, Appli |  |  |  |  |  |
| 12         | 4     | 9.5           | 2      | 2  | US-08-483-236-13  | Sequence 13, Appli |  |  |  |  |  |
| 13         | 4     | 9.5           | 2      | 4  | US-09-284-625-29  | Sequence 29, Appli |  |  |  |  |  |
| 14         | 4     | 9.5           | 2      | 4  | US-09-295-996B-58 | Sequence 58, Appli |  |  |  |  |  |
| 15         | 4     | 9.5           | 2      | 5  | PCT-US93-06625-1  | Sequence 1, Appli  |  |  |  |  |  |
| 16         | 4     | 9.5           | 3      | 1  | US-07-914-280-2   | Sequence 2, Appli  |  |  |  |  |  |
| 17         | 4     | 9.5           | 3      | 1  | US-07-791-213D-22 | Sequence 22, Appli |  |  |  |  |  |
| 18         | 4     | 9.5           | 3      | 1  | US-07-731-157A-14 | Sequence 14, Appli |  |  |  |  |  |
| 19         | 4     | 9.5           | 3      | 1  | US-08-127-904-6   | Sequence 6, Appli  |  |  |  |  |  |
| 20         | 4     | 9.5           | 3      | 1  | US-08-451-027-2   | Sequence 2, Appli  |  |  |  |  |  |
| 21         | 4     | 9.5           | 3      | 1  | US-08-202-178-8   | Sequence 8, Appli  |  |  |  |  |  |
| 22         | 4     | 9.5           | 3      | 1  | US-08-431-539-19  | Sequence 19, Appli |  |  |  |  |  |
| 23         | 4     | 9.5           | 3      | 1  | US-08-213-897A-13 | Sequence 13, Appli |  |  |  |  |  |
| 24         | 4     | 9.5           | 3      | 1  | US-08-713-897A-14 | Sequence 14, Appli |  |  |  |  |  |
| 25         | 4     | 9.5           | 3      | 1  | US-08-713-484-8   | Sequence 8, Appli  |  |  |  |  |  |
| 26         | 4     | 9.5           | 3      | 1  | US-08-670-136-2   | Sequence 2, Appli  |  |  |  |  |  |
| 27         | 4     | 9.5           | 3      | 1  | US-08-411-727-28  | Sequence 28, Appli |  |  |  |  |  |

ALIGNMENTS

RESULT 1  
US-07-914-280-1  
; Sequence 1, Application us/07914280  
; Patent No. 5304497  
; GENERAL INFORMATION:  
; APPLICANT: Boyd, Victoria L.  
; APPLICANT: Bozzini, Merilisa  
; APPLICANT: Gug, Piotr J.  
; TITLE OF INVENTION: Method of Forming N-Protected Amino Acid  
; TITLE OF INVENTION: Thionydantolins  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Law Offices of Peter Dehlinger  
; STREET: 350 Cambridge Avenue, Suite 300  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/07/914,280  
; APPLICATION NUMBER: 19920715  
; FILING DATE: 19920715  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fabian, Gary R.  
; REGISTRATION NUMBER: 33,875  
; REFERENCE/DOCKET NUMBER: 0550-0025  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: test peptide  
US-07-914-280-1

Query Match 9.5%; Score 4; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 1.8e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Sequence 21, Appli  
Sequence 25, Appli  
Sequence 12, Appli  
Sequence 44, Appli  
Sequence 51, Appli  
Sequence 4, Appli  
Sequence 9, Appli  
Sequence 10, Appli  
Sequence 22, Appli  
Sequence 4, Appli  
Sequence 5, Appli  
Sequence 11, Appli  
Sequence 44, Appli  
Sequence 51, Appli  
Sequence 21, Appli  
Sequence 25, Appli  
Sequence 12, Appli  
Sequence 13, Appli

QY 18 A 18  
Db 1 A 1

RESULT 2  
US-07-791-213D-23  
; Sequence 23 Application US/07791213D  
; Patent No. 5409895  
; GENERAL INFORMATION:  
; APPLICANT: MORISHITA, Hideaki  
; APPLICANT: KANAMORI, Toshinori  
; APPLICANT: NOBUHARA, Masahiro  
; TITLE OF INVENTION: POLYPEPTIDE, DNA FRAGMENT ENCODING THE  
; TITLE OF INVENTION: SAME AND PROCESS FOR PRODUCING THE SAME, AND ENZYME  
; TITLE OF INVENTION: INHIBITION PROCESS, DRUG COMPOSITION AND METHODS OF  
; TITLE OF INVENTION: TREATING USING THE SAME  
; NUMBER OF SEQUENCES: 108  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/791,213D  
; FILING DATE: 13-NOV-1991  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 2-306745  
; FILING DATE: 13-NOV-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Neuth, Donna M  
; REGISTRATION NUMBER: 36,607  
; REFERENCE/DOCKET NUMBER: 029650-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-07-791-213D-23  
Query Match 9.5%; Score 4; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 1.8e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 18 A 18  
Db 1 A 1

RESULT 3  
US-07-840-077A-5  
; Sequence 5, Application US/07840077A  
; Patent No. 5443816  
; GENERAL INFORMATION:  
; APPLICANT: Zamora, Paul O.  
; APPLICANT: Rhodes, Buck A.  
; TITLE OF INVENTION: Peptide-Metal Ion  
; TITLE OF INVENTION: Pharmaceutical Preparation and Method  
; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhomed Incorporated  
; STREET: 4261 Balloon Park  
; CITY: Albuquerque  
; STATE: NM  
; COUNTRY: U.S.A.  
; ZIP: 87109-5802  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb  
; MEDIUM TYPE: Storage  
; COMPUTER: IBM PC/XI/AT, IBM PS/2 or  
; COMPUTER: compatibles  
; OPERATING SYSTEM: PC-DOS or MS-DOS  
; SOFTWARE: WordPerfect 6.0a for Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/840,077A  
; FILING DATE: 20-FEB-1992  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/816,477  
; FILING DATE: 03-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Deborah A. Peacock  
; REGISTRATION NUMBER: 31,649  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (505) 242-9677  
; TELEFAX: (505) 243-2542  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-07-840-077A-5  
Query Match 9.5%; Score 4; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 1.8e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 18 A 18  
Db 2 A 2

RESULT 4  
US-08-122-510-10  
; Sequence 10, Application US/08122510  
; Patent No. 5464821  
; GENERAL INFORMATION:  
; APPLICANT: AASMUL-OLSEN, Stig  
; APPLICANT: WIDMER, Fred  
; APPLICANT: GAURI, Kailash  
; TITLE OF INVENTION: Small peptidic compounds useful for the  
; TITLE OF INVENTION: treatment of Glaucoma  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Allegretti & Witcoff, Ltd.  
; STREET: 10 S. Wacker Dr.  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/122,510  
; FILING DATE: 24-SEP-1993

CLASSIFICATION: 514  
PRIOR APPLICATION DATA: US 07/664,989  
FILING DATE: 01-MAR-1991  
APPLICATION NUMBER: 01-MAR-1991  
FILING DATE: 02-MAR-1990  
APPLICATION NUMBER: 02-MAR-1990  
FILING DATE: 02-SEP-1988  
APPLICATION NUMBER: 02-SEP-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Cooper, Iver P.  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: LEY-1  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..2  
OTHER INFORMATION: /note= "Where X is Benzylester  
US-08-122-510-10

Query Match 9.5% Score 4; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 1.8e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
Db 1 A 1

RESULT 5  
US-08-358-160-169  
Sequence 169, Application US/08358160  
Patent No. 5663143  
GENERAL INFORMATION:  
APPLICANT: LEY, Arthur C.  
APPLICANT: LADNER, Robert C.  
APPLICANT: GUTERMAN, Sonia K.  
APPLICANT: ROBERTS, Bruce L.  
APPLICANT: MARKLAND, William  
APPLICANT: KENT, Rachel B.  
TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ  
NUMBER OF SEQUENCES: 234  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W. Suite 300  
CITY: Washington  
STATE: District of Columbia  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/358,160  
FILING DATE: 16-DEC-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/133,031  
FILING DATE: 13-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/009,319

FILING DATE: 26-JAN-1993  
PRIOR APPLICATION DATA: US 07/664,989  
FILING DATE: 01-MAR-1991  
APPLICATION NUMBER: 01-MAR-1991  
FILING DATE: 02-MAR-1990  
APPLICATION NUMBER: 02-MAR-1990  
FILING DATE: 02-SEP-1988  
APPLICATION NUMBER: 02-SEP-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Cooper, Iver P.  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: LEY-1  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 169:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-358-160-169  
Query Match 9.5% Score 4; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 1.8e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 18 A 18  
Db 1 A 1  
RESULT 6  
US-08-358-160-170  
Sequence 170, Application US/08358160  
Patent No. 5663143  
GENERAL INFORMATION:  
APPLICANT: LEY, Arthur C.  
APPLICANT: LADNER, Robert C.  
APPLICANT: GUTERMAN, Sonia K.  
APPLICANT: ROBERTS, Bruce L.  
APPLICANT: MARKLAND, William  
APPLICANT: KENT, Rachel B.  
TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ  
NUMBER OF SEQUENCES: 234  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W. Suite 300  
CITY: Washington  
STATE: District of Columbia  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/358,160  
FILING DATE: 16-DEC-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/133,031  
FILING DATE: 13-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/009,319  
FILING DATE: 26-JAN-1993  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/664,989  
FILING DATE: 01-MAR-1991  
PRIOR APPLICATION DATA: US 07/487,063  
APPLICATION NUMBER:  
FILING DATE: 02-MAR-1990  
PRIOR APPLICATION DATA: US 07/240,160  
APPLICATION NUMBER: US 02-SEP-1988  
FILING DATE: 02-SEP-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Cooper, Iver P.  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: LEY=1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 170:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-358-160-170

Query Match 9.5%; Score 4; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 1.8e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0;

OY 18 A 18  
|  
Db 2 A 2

RESULT 7  
US-08-454-950-5  
; Sequence 5, Application US/08454950  
; Patent No. 5690905  
; GENERAL INFORMATION:  
; APPLICANT: Zamora, Paul O.  
; APPLICANT: Rhodes, Buck A.  
; TITLE OF INVENTION: Peptide-Metal Ion  
; TITLE OF INVENTION: Pharmaceutical Preparation and Method  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhomed Incorporated  
; STREET: 4261 Balloon Park  
; CITY: Albuquerque  
; STATE: NM  
; COUNTRY: U.S.A.  
; ZIP: 87109-5802  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb  
; MEDIUM TYPE: Storage  
; COMPUTER: IBM PC/XT/AT, IBM PS/2 or  
; COMPUTER: Compatibles  
; OPERATING SYSTEM: PC-DOS or MS-DOS  
; SOFTWARE: WordPerfect 6.0a for Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/454,950  
; FILING DATE: 31-MAY-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/840,077  
; FILING DATE: 20-FEB-1992  
; APPLICATION NUMBER: 07/816,477  
; FILING DATE: 03-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Deborah A. Peacock  
; REGISTRATION NUMBER: 31,649  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (505) 242-9677  
; TELEFAX: (505) 243-2542  
; INFORMATION FOR SEQ ID NO: 170:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-454-949-5

Query Match 9.5%; Score 4; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 1.8e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0;

INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-454-950-5  
Query Match 9.5%; Score 4; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 1.8e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0;  
OY 18 A 18  
|  
Db 2 A 2  
RESULT 8  
US-08-454-949-5  
; Sequence 5, Application US/08454949  
; Patent No. 5759516  
; GENERAL INFORMATION:  
; APPLICANT: Zamora, Paul O.  
; APPLICANT: Rhodes, Buck A.  
; TITLE OF INVENTION: Peptide-Metal Ion  
; TITLE OF INVENTION: Pharmaceutical Preparation and Method  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhomed Incorporated  
; STREET: 4261 Balloon Park  
; CITY: Albuquerque  
; STATE: NM  
; COUNTRY: U.S.A.  
; ZIP: 87109-5802  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb  
; MEDIUM TYPE: Storage  
; COMPUTER: IBM PC/XT/AT, IBM PS/2 or  
; COMPUTER: Compatibles  
; OPERATING SYSTEM: PC-DOS or MS-DOS  
; SOFTWARE: WordPerfect 6.0a for Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/454,949  
; FILING DATE: 31-MAY-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/840,077  
; FILING DATE: 20-FEB-1992  
; APPLICATION NUMBER: 07/816,477  
; FILING DATE: 03-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Deborah A. Peacock  
; REGISTRATION NUMBER: 31,649  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (505) 242-9677  
; TELEFAX: (505) 243-2542  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-454-949-5

Query Match 9.5%; Score 4; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 1.8e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0;



Qy 18 A 18  
Db 2 A 2

## RESULT 9

US-08-153-799-20  
; Sequence 20, Application US/08153799  
; Patent No. 5765883  
; GENERAL INFORMATION:  
; APPLICANT: Ballance, David J  
; APPLICANT: Goodey, Andrew R  
; TITLE OF INVENTION: Polypeptides  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: R Hain Swope, BOC Health Care Inc  
; STREET: 100 Mountain Avenue  
; CITY: Murray Hill  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07974  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/153,799  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/847975  
; FILING DATE: 06-MAR-1992  
; PRIOR APPLICATION DATA: GB 8909916.2  
; APPLICATION NUMBER: GB 8909916.2  
; FILING DATE: 29-APR-1989  
; PRIOR APPLICATION DATA: PCT/GB90/00650  
; APPLICATION NUMBER: PCT/GB90/00650  
; FILING DATE: 26-APR-1990  
; PRIOR APPLICATION DATA: US 07/775952  
; APPLICATION NUMBER: US 07/775952  
; FILING DATE: 29-OCT-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Swope, R Hain  
; REGISTRATION NUMBER: 24864  
; REFERENCE/DOCKET NUMBER: 92H832  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908) 665 2400  
; TELEFAX: (908) 771 6159  
; TELEX: 219484  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-153-799-20

Query Match 9.5%; Score 4; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 1.8e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
Db 2 A 2

## RESULT 10

US-08-293-150A-23  
; Sequence 23, Application US/08293150A  
; Patent No. 5792629  
; GENERAL INFORMATION:

; APPLICANT: MORISHITA, Hideaki  
; APPLICANT: KANAMORI, Toshinori  
; APPLICANT: NOBUHARA, Masahiro  
; TITLE OF INVENTION: POLYPEPTIDE, DNA FRAGMENT ENCODING THE  
; TITLE OF INVENTION: SAME AND PROCESS FOR PRODUCING THE SAME, AND ENZYME  
; TITLE OF INVENTION: INHIBITION PROCESS, DRUG COMPOSITION AND METHODS OF  
; TITLE OF INVENTION: TREATING USING THE SAME  
; NUMBER OF SEQUENCES: 110  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/293,150A  
; FILING DATE: 19-AUG-1994  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/791,213  
; FILING DATE: 13-NOV-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 2-306745  
; FILING DATE: 13-NOV-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meuth, Donna M.  
; REGISTRATION NUMBER: 36,607  
; REFERENCE/DOCKET NUMBER: 029650-049  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-293-150A-23

Query Match 9.5%; Score 4; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 1.8e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
Db 1 A 1

## RESULT 11

US-08-727-688-22  
; Sequence 22, Application US/08727688  
; Patent No. 5919638  
; GENERAL INFORMATION:  
; APPLICANT: Russell, John C.  
; TITLE OF INVENTION: Reagents and Methods for Detecting Prostate Tumors  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road D377/AP6D  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA: US/08/727,688  
APPLICATION NUMBER: 514  
FILING DATE: 1999-04-21  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 1  
FILING DATE: 1999-04-21  
ATTORNEY/AGENT INFORMATION:  
NAME: Porebski, Priscilla E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5967.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (847) 937-0378  
TELEFAX: (847) 938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 5919638e  
US-08-727-688-22

Query Match 9.5%; Score 4; DB 2; Length 2;  
Best Local Similarity 100.0%; Pred. No. 1.8e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0;  
Gaps 0;  
QY 18 A 18  
Db 1 A 1  
RESULT 12  
US-08-483-236-13  
Sequence 13, Application US/08483236  
Patent No. 5939385  
GENERAL INFORMATION:  
APPLICANT: Labroo, Virender  
APPLICANT: Busby, Sharon  
TITLE OF INVENTION: Transglutaminase Cross-Linkable  
POLYPEPTIDES AND METHODS RELATING THERETO  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ZymoGenetics, Inc.  
STREET: 1201 Eastlake Avenue East  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,236  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, Gary E.  
REGISTRATION NUMBER: 31-648  
REFERENCE/DOCKET NUMBER: 93-09c1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-442-6673  
TELEFAX: 206-442-6678  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
US-08-483-236-13

Query Match 9.5%; Score 4; DB 2; Length 2;  
Best Local Similarity 100.0%; Pred. No. 1.8e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0;  
Gaps 0;  
QY 18 A 18  
Db 1 A 1

RESULT 13  
US-09-284-625-29  
Sequence 29, Application US/09284625  
Patent No. 6207644  
GENERAL INFORMATION:  
APPLICANT: Luke, Richard William  
APPLICANT: Cotton, Richard  
TITLE OF INVENTION: Peptide Analogues Containing a 7-membered Lactam Ring  
FILE REFERENCE: 1991-169  
CURRENT APPLICATION NUMBER: US/09/284,625  
CURRENT FILING DATE: 1999-04-16  
PRIOR APPLICATION NUMBER: PCT/GB97/02837  
PRIOR FILING DATE: 1997-10-14  
PRIOR APPLICATION NUMBER: GB 96218367  
PRIOR FILING DATE: 1996-10-19  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 29  
LENGTH: 2  
TYPE: PPT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: MOD.RES  
LOCATION: (2)  
OTHER INFORMATION: Xaa= cyclohexylalanine  
US-09-284-625-29

Query Match 9.5%; Score 4; DB 4; Length 2;  
Best Local Similarity 100.0%; Pred. No. 1.8e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0;  
Gaps 0;  
QY 18 A 18  
Db 1 A 1

RESULT 14  
US-09-295-996B-58  
Sequence 58, Application US/09295996B  
Patent No. 6413530  
GENERAL INFORMATION:  
APPLICANT: Botovsky, Dov  
TITLE OF INVENTION: PESTICIDAL PEPTIDES  
FILE REFERENCE: UP-230  
CURRENT APPLICATION NUMBER: US/09/295,996B  
CURRENT FILING DATE: 1999-04-21  
NUMBER OF SEQ ID NOS: 60  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 58  
LENGTH: 2  
TYPE: PPT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: TMOF peptide  
US-09-295-996B-58

Query Match 9.5%; Score 4; DB 4; Length 2;  
Best Local Similarity 100.0%; Pred. No. 1.8e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0;  
Gaps 0;  
QY 18 A 18

Db 1 A 1

RESULT 15  
PCT-US93-06625-1  
: Sequence 1: Application PC/TUS9306625  
: GENERAL INFORMATION:  
: APPLICANT: Applied Biosystems, Inc.  
: TITLE OF INVENTION: METHOD OF FORMING N-PROTECTED AMINO ACID  
: NUMBER OF SEQUENCES: 14  
: CORRESPONDENCE ADDRESSES:  
: ADDRESSEE: Law Offices of Peter Dehlinger  
: STREET: 350 Cambridge Avenue, Suite 300  
: CITY: Palo Alto  
: STATE: CA  
: COUNTRY: USA  
: ZIP: 94306  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC Compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patent Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: PCT/US93/06625  
: FILING DATE: 19930714  
: CLASSIFICATION:  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/914,280  
: FILING DATE: 15-JUL-1992  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Powers, Vincent M.  
: REGISTRATION NUMBER: 36,246  
: REFERENCE/DOCKET NUMBER: 0550-0025.41  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (415) 324-0880  
: TELEFAX: (415) 324-0960  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 2 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: peptide  
: ORIGINAL SOURCE:  
: INDIVIDUAL ISOLATE: synthetic peptide  
PCT-US93-06625-1

Query Match 9.5% Score 4; DB 5; Length 2;  
Best Local Similarity 100.0%; Pred No. 1.8e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18  
Db 1 A 1

Search completed: January 8, 2003, 10:59:32  
Job time : 12.0948 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2003, 10:55:49 ; Search time 7.06034 Seconds  
(without alignments)  
107.168 Million cell updates/sec

Title: US-09-003-869-5  
Perfect score: 42  
Sequence: 1 xx 39

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 118974 seqs, 19401057 residues

al number of hits satisfying chosen parameters: 118974

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA.\*

1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description       |
|------------|-------|-------------|--------|-------|-------------------|
| 1          | 4     | 9.5         | 2      | 10    | US-09-816-737-14  |
| 2          | 4     | 9.5         | 2      | 10    | US-09-821-883-14  |
| 3          | 4     | 9.5         | 2      | 10    | US-09-982-172-68  |
| 4          | 4     | 9.5         | 2      | 10    | US-09-982-172-116 |
| 5          | 4     | 9.5         | 2      | 10    | US-09-982-172-141 |
| 6          | 4     | 9.5         | 3      | 9     | US-10-003-035-13  |
| 7          | 4     | 9.5         | 3      | 9     | US-09-861-688-5   |
| 8          | 4     | 9.5         | 3      | 9     | US-10-024-860-3   |
| 9          | 4     | 9.5         | 3      | 9     | US-10-099-895-45  |
| 10         | 4     | 9.5         | 3      | 10    | US-09-113-696B-36 |
| 11         | 4     | 9.5         | 3      | 10    | US-09-816-737-12  |
| 12         | 4     | 9.5         | 3      | 10    | US-09-816-737-13  |
| 13         | 4     | 9.5         | 3      | 10    | US-09-821-883-13  |
| 14         | 4     | 9.5         | 3      | 10    | US-09-853-918-4   |
| 15         | 4     | 9.5         | 3      | 10    | US-09-853-918-7   |
| 16         | 4     | 9.5         | 3      | 10    | US-09-972-475-20  |
| 17         | 4     | 9.5         | 3      | 10    | US-09-326-447-12  |
| 18         | 4     | 9.5         | 3      | 10    | US-09-967-003-4   |
| 19         | 4     | 9.5         | 3      | 10    | US-09-984-056-25  |

20 4 9.5 3 10 US-09-982-172-58 Sequence 58, Appl  
21 4 9.5 3 10 US-09-982-172-130 Sequence 130, Appl  
22 4 9.5 3 10 US-09-984-057-25 Sequence 25, Appl  
23 4 9.5 4 8 US-08-450-842-78 Sequence 78, Appl  
24 4 9.5 4 8 US-08-450-842-84 Sequence 84, Appl  
25 4 9.5 4 8 US-08-450-842-86 Sequence 86, Appl  
26 4 9.5 4 8 US-08-946-839-1 Sequence 1, Appl  
27 4 9.5 4 8 US-08-946-839-2 Sequence 2, Appl  
28 4 9.5 4 8 US-08-864-392B-6 Sequence 6, Appl  
29 4 9.5 4 8 US-08-484-409-3 Sequence 3, Appl  
30 4 9.5 4 8 US-08-484-409-4 Sequence 4, Appl  
31 4 9.5 4 8 US-08-484-409-10 Sequence 10, Appl  
32 4 9.5 4 8 US-08-484-409-21 Sequence 21, Appl  
33 4 9.5 4 8 US-08-484-409-27 Sequence 27, Appl  
34 4 9.5 4 8 US-08-484-409-28 Sequence 28, Appl  
35 4 9.5 4 8 US-08-484-409-33 Sequence 33, Appl  
36 4 9.5 4 8 US-08-981-824-43 Sequence 43, Appl  
37 4 9.5 4 8 US-08-424-550B-563 Sequence 563, Appl  
38 4 9.5 4 9 US-09-742-096-20 Sequence 20, Appl  
39 4 9.5 4 9 US-09-742-096-27 Sequence 27, Appl  
40 4 9.5 4 9 US-09-742-096-28 Sequence 28, Appl  
41 4 9.5 4 9 US-09-946-605-8 Sequence 8, Appl  
42 4 9.5 4 9 US-09-946-605-11 Sequence 11, Appl  
43 4 9.5 4 9 US-09-964-114-3 Sequence 3, Appl  
44 4 9.5 4 9 US-09-964-114-4 Sequence 4, Appl  
45 4 9.5 4 9 US-09-964-114-6 Sequence 6, Appl

#### ALIGNMENTS

RESULT 1  
US-09-816-737-14  
; Sequence 14, Application US/09816737  
; Patent No. US20020037853A1  
; GENERAL INFORMATION:  
; APPLICANT: Bhatnagar, Rajendra S.  
; TITLE OF INVENTION: "Synthetic Compounds and Compositions with Enhanced Cell Binding"  
; FILE REFERENCE: 06510223CON2  
; CURRENT APPLICATION NUMBER: US/09/816,737  
; CURRENT FILING DATE: 2001-03-23  
; PRIOR APPLICATION NUMBER: 09/328,347  
; PRIOR FILING DATE: 1999-06-08  
; PRIOR APPLICATION NUMBER: 08/859,610  
; PRIOR FILING DATE: 1997-05-20  
; PRIOR APPLICATION NUMBER: 08/278,878  
; PRIOR FILING DATE: 1994-07-22  
; PRIOR APPLICATION NUMBER: 07/804,782  
; PRIOR FILING DATE: 1991-12-09  
; PRIOR APPLICATION NUMBER: 07/393,621  
; PRIOR FILING DATE: 1989-08-14  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 2  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic peptide  
; NAME/KEY: ACETYLATION  
; LOCATION: 1  
; OTHER INFORMATION: N-acetyl Isoleucine  
; NAME/KEY: METHYLATION  
; LOCATION: 2  
; OTHER INFORMATION: alanyl-methylamide  
US-09-816-737-14

Query Match 9.5%; Score 4; DB 10; Length 2;  
Best Local Similarity 100.0%; Pred. No. 8.8e+04;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18



```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 3
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: GAG site
US-10-003-035-13

Query Match          9.5%: Score 4; DB 9; Length 3;
Best Local Similarity 100.0%; Pred. No. 8.8e+04;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
Db 2 A 2

RESULT 7
US-09-861-688-5
; Sequence 5, Application US/09861688
; Patent No. US20020173460A1
; GENERAL INFORMATION:
; APPLICANT: Clargen, Inc. & NIH
; TITLE OF INVENTION: Use of Recombinant Human Uteroglobulin in Treatment of
; TITLE OF INVENTION: Inflammatory and
; TITLE OF INVENTION: Fibrotic Conditions
; FILE REFERENCE: 116142/2
; CURRENT APPLICATION NUMBER: US/09/861,688
; CURRENT FILING DATE: 2001-05-21
; PRIOR FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 3
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer sequence
US-09-861-688-5

Query Match          9.5%: Score 4; DB 9; Length 3;
Best Local Similarity 100.0%; Pred. No. 8.8e+04;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
Db 2 A 2

RESULT 8
US-10-024-860-3
; Sequence 3, Application US/10024860
; Patent No. US20020172692A1
; GENERAL INFORMATION:
; APPLICANT: Cohen, Joseph
; TITLE OF INVENTION: Vaccine Composition Against Malaria
; FILE REFERENCE: B45088C2
; CURRENT APPLICATION NUMBER: US/10/024,860
; CURRENT FILING DATE: 2001-12-18
; PRIOR FILING DATE: 2001-04-05
; PRIOR FILING DATE: 1999-01-26
; PRIOR FILING DATE: 1999-01-26
; PRIOR FILING DATE: 1996-08-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 3
; TYPE: PRT
; ORGANISM: Hepatitis B

US-10-024-860-3
; Sequence 36, Application US/09113696B
; Patent No. US20020010134A1
; GENERAL INFORMATION:
; APPLICANT: Bhatnagar, Rajendra S.
; APPLICANT: Qian, Jing Jing
; APPLICANT: Gough, Craig
; TITLE OF INVENTION: PEPTIDE COMPOSITIONS MIMICKING TGF-BETA
; FILE REFERENCE: 6510-21SCIP2
; CURRENT APPLICATION NUMBER: US/09/113,696B
; CURRENT FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: 08/742,256
; PRIOR FILING DATE: 1996-10-31
; PRIOR APPLICATION NUMBER: 08/431,954
; PRIOR FILING DATE: 1995-05-01
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 3
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Collagen receptor ligands
; NAME/KEY: MOD.RES
; LOCATION: (3)...(3)
; OTHER INFORMATION: bala
US-09-113-696B-36

US-10-024-860-3
; Sequence 45, Application US/10099895
; Patent No. US2002017166A1
; GENERAL INFORMATION:
; APPLICANT: BERNDT, Michael C
; APPLICANT: STOMSKI, Frank C
; APPLICANT: LOPEZ, Angel F
; APPLICANT: GUTHRIE, Mark A
; TITLE OF INVENTION: A BINDING MOTIF OF A RECEPTOR
; FILE REFERENCE: 3991/0K379
; CURRENT APPLICATION NUMBER: US/10/099,895
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: PCT/A000/01118
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45
; LENGTH: 3
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-099-895-45

Query Match          9.5%: Score 4; DB 9; Length 3;
Best Local Similarity 100.0%; Pred. No. 8.8e+04;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
Db 1 A 1

RESULT 10
US-09-113-696B-36
; Sequence 36, Application US/09113696B
; Patent No. US20020010134A1
; GENERAL INFORMATION:
; APPLICANT: Bhatnagar, Rajendra S.
; APPLICANT: Qian, Jing Jing
; APPLICANT: Gough, Craig
; TITLE OF INVENTION: PEPTIDE COMPOSITIONS MIMICKING TGF-BETA
; FILE REFERENCE: 6510-21SCIP2
; CURRENT APPLICATION NUMBER: US/09/113,696B
; CURRENT FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: 08/742,256
; PRIOR FILING DATE: 1996-10-31
; PRIOR APPLICATION NUMBER: 08/431,954
; PRIOR FILING DATE: 1995-05-01
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 3
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Collagen receptor ligands
; NAME/KEY: MOD.RES
; LOCATION: (3)...(3)
; OTHER INFORMATION: bala
US-09-113-696B-36
```

Query Match 9.5%; Score 4; DB 10; Length 3;  
Best Local Similarity 100.0%; Pred. No. 8.8e+04;  
Matches 1; Conservative 0; Mismatches 0; Indels 0;

QY 18 A 18  
|  
Db 2 A 2

RESULT 11  
US-09-816-737-12  
; Sequence 12, Application US/09816737  
; Patent No. US20020037853A1  
; GENERAL INFORMATION:  
; APPLICANT: Bhattacharya, Rajendra S.  
; TITLE OF INVENTION: "Synthetic Compounds and Compositions  
; with Enhanced Cell Binding"  
; FILE REFERENCE: 06510223CON2  
; CURRENT APPLICATION NUMBER: US/09/816,737  
; CURRENT FILING DATE: 2001-03-23  
; PRIOR APPLICATION NUMBER: 09/328,347  
; PRIOR FILING DATE: 1999-06-08  
; PRIOR APPLICATION NUMBER: 08/859,610  
; PRIOR FILING DATE: 1997-05-20  
; PRIOR APPLICATION NUMBER: 08/278,878  
; PRIOR FILING DATE: 1994-07-22  
; PRIOR APPLICATION NUMBER: 07/804,782  
; PRIOR FILING DATE: 1991-12-09  
; PRIOR APPLICATION NUMBER: 07/393,621  
; PRIOR FILING DATE: 1989-08-14  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic peptide  
; NAME/KEY: ACETYLATION  
; LOCATION: 1  
; OTHER INFORMATION: N-acetyl Isoleucine  
US-09-816-737-12

Query Match 9.5%; Score 4; DB 10; Length 3;  
Best Local Similarity 100.0%; Pred. No. 8.8e+04;  
Matches 1; Conservative 0; Mismatches 0; Indels 0;

18 A 18  
|  
2 A 2

RESULT 12  
US-09-816-737-13  
; Sequence 13, Application US/09816737  
; Patent No. US20020037853A1  
; GENERAL INFORMATION:  
; APPLICANT: Bhattacharya, Rajendra S.  
; TITLE OF INVENTION: "Synthetic Compounds and Compositions  
; with Enhanced Cell Binding"  
; FILE REFERENCE: 06510223CON2  
; CURRENT APPLICATION NUMBER: US/09/816,737  
; CURRENT FILING DATE: 2001-03-23  
; PRIOR APPLICATION NUMBER: 09/328,347  
; PRIOR FILING DATE: 1999-06-08  
; PRIOR APPLICATION NUMBER: 08/859,610  
; PRIOR FILING DATE: 1997-05-20  
; PRIOR APPLICATION NUMBER: 08/278,878  
; PRIOR FILING DATE: 1994-07-22  
; PRIOR APPLICATION NUMBER: 07/804,782  
; PRIOR FILING DATE: 1991-12-09  
; PRIOR APPLICATION NUMBER: 07/393,621  
; PRIOR FILING DATE: 1989-08-14

; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic peptide  
; NAME/KEY: MOD\_RES  
; LOCATION: 3  
; OTHER INFORMATION: bala  
US-09-816-737-13

Query Match 9.5%; Score 4; DB 10; Length 3;  
Best Local Similarity 100.0%; Pred. No. 8.8e+04;  
Matches 1; Conservative 0; Mismatches 0; Indels 0;

QY 18 A 18  
|  
Db 2 A 2

RESULT 13  
US-09-821-883-13  
; Sequence 13, Application US/09821883  
; Patent No. US20020061310A1  
; GENERAL INFORMATION:  
; APPLICANT: Laus, Reiner  
; APPLICANT: Vidovic, Damir  
; APPLICANT: Graddis, Thomas  
; TITLE OF INVENTION: Compositions and Methods for Dendritic  
; CELL-BASED IMMUNOTHERAPY  
; FILE REFERENCE: 7636-0022.30  
; CURRENT APPLICATION NUMBER: US/09/821,883  
; CURRENT FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: US 60/193,504  
; PRIOR FILING DATE: 2000-03-30  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-821-883-13

Query Match 9.5%; Score 4; DB 10; Length 3;  
Best Local Similarity 100.0%; Pred. No. 8.8e+04;  
Matches 1; Conservative 0; Mismatches 0; Indels 0;

QY 18 A 18  
|  
Db 2 A 2

RESULT 14  
US-09-853-918-4  
; Sequence 4, Application US/09853918  
; Patent No. US20020068346A1  
; GENERAL INFORMATION:  
; APPLICANT: Krystek, Stanley R.  
; APPLICANT: Sheriff, Steven  
; APPLICANT: Witmer, Mark R.  
; APPLICANT: Hollenbaugh, Diane L.  
; APPLICANT: Yan, Ning  
; APPLICANT: Mouravieff, Julie E.  
; APPLICANT: Einspahr, Howard M.  
; APPLICANT: Kish, Kevin  
; TITLE OF INVENTION: MODIFIED INOSINE 5'-MONOPHOSPHATE DEHYDROGENASE  
; POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: DB24NP  
; CURRENT APPLICATION NUMBER: US/09/853,918  
; CURRENT FILING DATE: 2001-05-10  
; PRIOR APPLICATION NUMBER: 60/203,448



; PRIOR FILING DATE: 2000-05-10  
; NUMBER OF SEQ ID NOS: 65  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-853-918-4

Query Match 9.5%; Score 4; DB 10; Length 3;  
Best Local Similarity 100.0%; Pred. No. 8.8e+04;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
|  
Db 2 A 2

RESULT 15

US-09-853-918-7  
Sequence 7, Application US/09853918  
Patent No. US20020068346A1  
; GENERAL INFORMATION:  
; APPLICANT: Krystek, Stanley R.  
; APPLICANT: Sheriff, Steven  
; APPLICANT: Witmer, Mark R.  
; APPLICANT: Hollenbaugh, Diane L.  
; APPLICANT: Yan, Ning  
; APPLICANT: Mouravieff, Julie E.  
; APPLICANT: Einspahr, Howard M.  
; APPLICANT: Kish, Kevin  
; TITLE OF INVENTION: MODIFIED INOSINE 5'-MONOPHOSPHATE DEHYDROGENASE  
; FILE OF INVENTION: POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: DB24NP  
; CURRENT APPLICATION NUMBER: US/09/853,918  
; CURRENT FILING DATE: 2001-05-10  
; PRIOR APPLICATION NUMBER: 60/203,448  
; PRIOR FILING DATE: 2000-05-10  
; NUMBER OF SEQ ID NOS: 65  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-853-918-7

Query Match 9.5%; Score 4; DB 10; Length 3;  
Best Local Similarity 100.0%; Pred. No. 8.8e+04;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
|  
Db 1 A 1

Search completed: January 8, 2003, 11:01:14  
Job time : 7.06034 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2003, 10:58:35 ; Search time 216.517 Seconds  
(without alignments)  
116.132 Million cell updates/sec

Title: US-09-003-869-5  
Perfect score: 42  
Sequence: 1 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 39

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 4569144 seqs, 644733110 residues 4569144  
al number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

| Database : |  |  |  | Pending Patents AA.Main.* |  |  |  |
|------------|--|--|--|---------------------------|--|--|--|
| 1:         | /cgn2_6/ptodata/1/paa/US06_COMB.pep.*  |  |  |                           |  |  |  |
| 2:         | /cgn2_6/ptodata/1/paa/US06_COMB.pep.*  |  |  |                           |  |  |  |
| 3:         | /cgn2_6/ptodata/1/paa/US07_COMB.pep.*  |  |  |                           |  |  |  |
| 4:         | /cgn2_6/ptodata/1/paa/US08_COMB.pep.*  |  |  |                           |  |  |  |
| 5:         | /cgn2_6/ptodata/1/paa/US08_COMB.pep.*  |  |  |                           |  |  |  |
| 6:         | /cgn2_6/ptodata/1/paa/US08_COMB.pep.*  |  |  |                           |  |  |  |
| 7:         | /cgn2_6/ptodata/1/paa/US08_COMB.pep.*  |  |  |                           |  |  |  |
| 8:         | /cgn2_6/ptodata/1/paa/US08_COMB.pep.*  |  |  |                           |  |  |  |
| 9:         | /cgn2_6/ptodata/1/paa/US08_COMB.pep.*  |  |  |                           |  |  |  |
| 10:        | /cgn2_6/ptodata/1/paa/US08_COMB.pep.*  |  |  |                           |  |  |  |
| 11:        | /cgn2_6/ptodata/1/paa/US08_COMB.pep.*  |  |  |                           |  |  |  |
| 12:        | /cgn2_6/ptodata/1/paa/US08_COMB.pep.*  |  |  |                           |  |  |  |
| 13:        | /cgn2_6/ptodata/1/paa/US08_COMB.pep.*  |  |  |                           |  |  |  |
| 14:        | /cgn2_6/ptodata/1/paa/US09_COMB.pep.*  |  |  |                           |  |  |  |
| 15:        | /cgn2_6/ptodata/1/paa/US09_COMB.pep.*  |  |  |                           |  |  |  |
| 16:        | /cgn2_6/ptodata/1/paa/US09_COMB.pep.*  |  |  |                           |  |  |  |
| 17:        | /cgn2_6/ptodata/1/paa/US09_COMB.pep.*  |  |  |                           |  |  |  |
| 18:        | /cgn2_6/ptodata/1/paa/US09_COMB.pep.*  |  |  |                           |  |  |  |
| 19:        | /cgn2_6/ptodata/1/paa/US09_COMB.pep.*  |  |  |                           |  |  |  |
| 20:        | /cgn2_6/ptodata/1/paa/US09_COMB.pep.*  |  |  |                           |  |  |  |
| 21:        | /cgn2_6/ptodata/1/paa/US09_COMB.pep.*  |  |  |                           |  |  |  |
| 22:        | /cgn2_6/ptodata/1/paa/US09_COMB.pep.*  |  |  |                           |  |  |  |
| 23:        | /cgn2_6/ptodata/1/paa/US09_COMB.pep.*  |  |  |                           |  |  |  |
| 24:        | /cgn2_6/ptodata/1/paa/US100_COMB.pep.* |  |  |                           |  |  |  |
| 25:        | /cgn2_6/ptodata/1/paa/US100_COMB.pep.* |  |  |                           |  |  |  |
| 26:        | /cgn2_6/ptodata/1/paa/US102_COMB.pep.* |  |  |                           |  |  |  |
| 27:        | /cgn2_6/ptodata/1/paa/US60_COMB.pep.*  |  |  |                           |  |  |  |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Score | Length | ID | Description         |
|------------|-------------|-------|--------|----|---------------------|
| 1          | 4           | 9.5   | 2      | 1  | PCT-US01-01786A-264 |
| 2          | 4           | 9.5   | 2      | 1  | PCT-US02-05064-322  |
| 3          | 4           | 9.5   | 2      | 1  | PCT-US02-09188-444  |
| 4          | 4           | 9.5   | 2      | 1  | PCT-US02-09239-129  |
| 5          | 4           | 9.5   | 2      | 1  | PCT-US02-09257-295  |
| 6          | 4           | 9.5   | 2      | 1  | PCT-US02-09370-454  |

|    |   |     |   |    |                     |                    |
|----|---|-----|---|----|---------------------|--------------------|
| 7  | 4 | 9.5 | 2 | 1  | PCT-US02-09922-279  | Sequence 279, App  |
| 8  | 4 | 9.5 | 2 | 6  | US-08-222-626A-104  | Sequence 104, App  |
| 9  | 4 | 9.5 | 2 | 6  | US-08-222-626A-118  | Sequence 118, App  |
| 10 | 4 | 9.5 | 2 | 6  | US-08-293-150-23    | Sequence 23, Appl  |
| 11 | 4 | 9.5 | 2 | 8  | US-08-476-075-1     | Sequence 1, Appl   |
| 12 | 4 | 9.5 | 2 | 9  | US-08-564-222-1     | Sequence 1, Appl   |
| 13 | 4 | 9.5 | 2 | 14 | US-09-077-574A-23   | Sequence 23, Appl  |
| 14 | 4 | 9.5 | 2 | 17 | US-09-302-905-13    | Sequence 13, Appl  |
| 15 | 4 | 9.5 | 2 | 17 | US-09-302-905A-13   | Sequence 40, Appl  |
| 16 | 4 | 9.5 | 2 | 17 | US-09-317-702-40    | Sequence 37, Appl  |
| 17 | 4 | 9.5 | 2 | 21 | US-09-762-370-37    | Sequence 14, Appl  |
| 18 | 4 | 9.5 | 2 | 22 | US-09-816-737-14    | Sequence 14, Appl  |
| 19 | 4 | 9.5 | 2 | 22 | US-09-821-883-14    | Sequence 77, Appl  |
| 20 | 4 | 9.5 | 2 | 23 | US-09-920-306-77    | Sequence 547, App  |
| 21 | 4 | 9.5 | 2 | 23 | US-09-950-082-547   | Sequence 68, Appl  |
| 22 | 4 | 9.5 | 2 | 23 | US-09-982-172-68    | Sequence 116, App  |
| 23 | 4 | 9.5 | 2 | 23 | US-09-982-172-116   | Sequence 141, App  |
| 24 | 4 | 9.5 | 2 | 25 | US-10-105-299-3144  | Sequence 3144, App |
| 25 | 4 | 9.5 | 2 | 25 | US-10-137-867-386   | Sequence 386, App  |
| 26 | 4 | 9.5 | 2 | 27 | US-60-173-464-1365  | Sequence 1365, App |
| 27 | 4 | 9.5 | 2 | 27 | US-60-191-637-28343 | Sequence 28343, A  |
| 28 | 4 | 9.5 | 2 | 27 | US-60-191-681-23152 | Sequence 23152, A  |
| 29 | 4 | 9.5 | 3 | 1  | PCT-US00-06059-61   | Sequence 61, Appl  |
| 30 | 4 | 9.5 | 3 | 1  | PCT-US00-06062-61   | Sequence 11, Appl  |
| 31 | 4 | 9.5 | 3 | 1  | PCT-US00-08879-11   | Sequence 12, Appl  |
| 32 | 4 | 9.5 | 3 | 1  | PCT-US00-08879-12   | Sequence 15, Appl  |
| 33 | 4 | 9.5 | 3 | 1  | PCT-US00-08879-14   | Sequence 17, Appl  |
| 34 | 4 | 9.5 | 3 | 1  | PCT-US00-08879-15   | Sequence 20, Appl  |
| 35 | 4 | 9.5 | 3 | 1  | PCT-US00-08879-20   | Sequence 25, Appl  |
| 36 | 4 | 9.5 | 3 | 1  | PCT-US00-08879-25   | Sequence 26, Appl  |
| 37 | 4 | 9.5 | 3 | 1  | PCT-US00-08879-26   | Sequence 29, Appl  |
| 38 | 4 | 9.5 | 3 | 1  | PCT-US00-08879-29   | Sequence 31, Appl  |
| 39 | 4 | 9.5 | 3 | 1  | PCT-US00-08879-31   | Sequence 11, Appl  |
| 40 | 4 | 9.5 | 3 | 1  | PCT-US00-08879A-11  | Sequence 12, Appl  |
| 41 | 4 | 9.5 | 3 | 1  | PCT-US00-08879A-12  | Sequence 14, Appl  |
| 42 | 4 | 9.5 | 3 | 1  | PCT-US00-08879A-14  | Sequence 15, Appl  |
| 43 | 4 | 9.5 | 3 | 1  | PCT-US00-08879A-15  |                    |
| 44 | 4 | 9.5 | 3 | 1  |                     |                    |
| 45 | 4 | 9.5 | 3 | 1  |                     |                    |

ALIGNMENTS

RESULT 1  
PCT-US01-01786A-264  
; Sequence 264, Application PC/TUS0101786A  
; GENERAL INFORMATION:  
; APPLICANT: SMITHKLINE BEECHAM CORPORATION  
; APPLICANT: SMITHKLINE BEECHAM p.l.c.  
; TITLE OF INVENTION: METHODS AND REAGENTS FOR PERFORMING  
; FILE REFERENCE: ANTIMICROBIAL COMPOUND SCREENING  
; FILE REFERENCE: GM50067  
; CURRENT APPLICATION NUMBER: PCT/US01/01786A  
; CURRENT FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/176,870  
; PRIOR FILING DATE: 2000-01-18  
; NUMBER OF SEQ ID NOS: 403  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 264  
; LENGTH: 2  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
PCT-US01-01786A-264

|                       |         |              |          |                         |
|-----------------------|---------|--------------|----------|-------------------------|
| Query Match           | 9.5%    | Score 4:     | DB 1:    | Length 2:               |
| Best Local Similarity | 100.0%  | Pred. No.    | 3.4e+06; |                         |
| Matches               | 1;      | Conservative | 0;       | Mismatches 0; Indels 0; |
| Qy                    | 18 A 18 |              |          |                         |
| Db                    | 1 A 1   |              |          |                         |

```
RESULT 2
PCT-US02-05064-322
; Sequence 322, Application PC/TUS0205064
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 83 Human Secreted Proteins
; FILE REFERENCE: PS735PCT
; CURRENT FILING DATE: 2002-03-28
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/270,658
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/304,444
; PRIOR FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 445
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 322
; LENGTH: 2
; TYPE: PRT
; ORGANISM: Homo sapiens
; US02-05064-322

Query Match          9.5%; Score 4; DB 1; Length 2;
Best Local Similarity 100.0%; Pred. No. 3.4e+06;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
Db 2 A 2

RESULT 3
PCT-US02-09188-444
; Sequence 444, Application PC/TUS0209188
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS952PCT
; CURRENT FILING DATE: 2002-03-26
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; NUMBER OF SEQ ID NOS: 1732
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 444
; LENGTH: 2
; TYPE: PRT
; ORGANISM: Homo sapiens
; PCT-US02-09188-444

Query Match          9.5%; Score 4; DB 1; Length 2;
Best Local Similarity 100.0%; Pred. No. 3.4e+06;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
Db 2 A 2

RESULT 4
PCT-US02-09239-129
; Sequence 129, Application PC/TUS0209239
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS953PCT
; CURRENT FILING DATE: 2002-03-26
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: PCT/US02/09239
; NUMBER OF SEQ ID NOS: 1834
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 129
; LENGTH: 2
; TYPE: PRT
; ORGANISM: Homo sapiens
; PCT-US02-09239-129

Query Match          9.5%; Score 4; DB 1; Length 2;
Best Local Similarity 100.0%; Pred. No. 3.4e+06;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
Db 2 A 2
```

```
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 380
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 129
; LENGTH: 2
; TYPE: PRT
; ORGANISM: Homo sapiens
; PCT-US02-09239-129

Query Match          9.5%; Score 4; DB 1; Length 2;
Best Local Similarity 100.0%; Pred. No. 3.4e+06;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
Db 2 A 2

RESULT 5
PCT-US02-09257-295
; Sequence 295, Application PC/TUS0209257
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS957PCT
; CURRENT FILING DATE: 2002-03-26
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; NUMBER OF SEQ ID NOS: 994
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 295
; LENGTH: 2
; TYPE: PRT
; ORGANISM: Homo sapiens
; PCT-US02-09257-295

Query Match          9.5%; Score 4; DB 1; Length 2;
Best Local Similarity 100.0%; Pred. No. 3.4e+06;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
Db 2 A 2

RESULT 6
PCT-US02-09370-454
; Sequence 454, Application PC/TUS0209370
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS954PCT
; CURRENT FILING DATE: 2002-03-26
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 60/278,650
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 09/950,082
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 09/950,083
; NUMBER OF SEQ ID NOS: 1834
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 454
; LENGTH: 2
; TYPE: PRT
; ORGANISM: Homo sapiens
; PCT-US02-09370-454

Query Match          9.5%; Score 4; DB 1; Length 2;
Best Local Similarity 100.0%; Pred. No. 3.4e+06;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
Db 2 A 2
```

TYPE: PRT  
ORGANISM: Homo sapiens  
PCT-US02-09370-454

Query Match 9.5%; Score 4; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 3.4e+06;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18  
DB 2 A 2

RESULT 7  
PCT-US02-09922-279  
Sequence 279, Application PC/TUS0209922  
GENERAL INFORMATION:  
APPLICANT: Human Genome Sciences, Inc.  
TITLE OF INVENTION: Human Secreted Proteins  
FILE REFERENCE: PS955PCT  
CURRENT APPLICATION NUMBER: PCT/US02/09922  
CURRENT FILING DATE: 2002-03-26  
PRIOR APPLICATION NUMBER: US 60/278,650  
PRIOR FILING DATE: 2001-03-27  
PRIOR APPLICATION NUMBER: US 09/950,082  
PRIOR FILING DATE: 2001-09-12  
PRIOR APPLICATION NUMBER: US 09/950,083  
PRIOR FILING DATE: 2001-09-12  
NUMBER OF SEQ ID NOS: 117  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 279  
LENGTH: 2  
TYPE: PRT  
ORGANISM: Homo sapiens  
PCT-US02-09922-279

Query Match 9.5%; Score 4; DB 1; Length 2;  
Best Local Similarity 100.0%; Pred. No. 3.4e+06;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18  
DB 2 A 2

RESULT 8  
US-08-222-626A-104  
Sequence 104, Application US/08222626A  
GENERAL INFORMATION:  
APPLICANT: Masuda, Esteban S.  
APPLICANT: Tokumitsu, Hiroshi  
TITLE OF INVENTION: PURIFIED COMPONENTS OF MAMMALIAN  
TITLE OF INVENTION: TRANSCRIPTION REGULATION COMPLEXES, AND ANALOGS  
NUMBER OF SEQUENCES: 183  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DNAX Research Institute  
STREET: 901 California Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/222,626A  
FILING DATE: 04-APR-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,061

FILING DATE: 05-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/113,971  
FILING DATE: 30-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/088,998  
FILING DATE: 30-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/088,483  
FILING DATE: 06-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Ching, Edwin P.  
REGISTRATION NUMBER: 34,090  
REFERENCE/DOCKET NUMBER: DX0392K3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-852-9196  
TELEFAX: 415-496-1200  
INFORMATION FOR SEQ ID NO: 104:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLSCULE TYPE: peptide  
US-08-222-626A-104

Query Match 9.5%; Score 4; DB 6; Length 2;  
Best Local Similarity 100.0%; Pred. No. 3.4e+06;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18  
DB 1 A 1

RESULT 9  
US-08-222-626A-118  
Sequence 118, Application US/08222626A  
GENERAL INFORMATION:  
APPLICANT: Aral, Naoko  
APPLICANT: Masuda, Esteban S.  
APPLICANT: Tokumitsu, Hiroshi  
TITLE OF INVENTION: PURIFIED COMPONENTS OF MAMMALIAN  
TITLE OF INVENTION: TRANSCRIPTION REGULATION COMPLEXES, AND ANALOGS  
NUMBER OF SEQUENCES: 183  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DNAX Research Institute  
STREET: 901 California Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/222,626A  
FILING DATE: 04-APR-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,061  
FILING DATE: 05-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/113,971  
FILING DATE: 30-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/088,998  
FILING DATE: 30-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/088,483  
FILING DATE: 06-JUL-1993

ATTORNEY/AGENT INFORMATION:  
NAME: CHING, EDWIN P.  
REGISTRATION NUMBER: 34,090  
REFERENCE/DOCKET NUMBER: DX0392K3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-852-9196  
TELEFAX: 415-496-1200  
INFORMATION FOR SEQ ID NO: 118:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-223-626A-118

Query Match 9.5%; Score 4; DB 6; Length 2;  
Best Local Similarity 100.0%; Pred. No. 3.4e+06;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

18 A 18

QY 1 A 1

## RESULT 10

US-08-293-150-23  
SEQUENCE 23, Application US/08293150  
GENERAL INFORMATION:  
APPLICANT: MORISHITA, Hideaki  
APPLICANT: KANAMORI, Toshinori  
APPLICANT: NOBUHARA, Masahiro  
TITLE OF INVENTION: POLYPEPTIDE, DNA FRAGMENT ENCODING THE  
TITLE OF INVENTION: SAME AND PROCESS FOR PRODUCING THE SAME, AND ENZYME  
TITLE OF INVENTION: INHIBITION PROCESS, DRUG COMPOSITION AND METHODS OF  
TITLE OF INVENTION: TREATING USING THE SAME  
NUMBER OF SEQUENCES: 108  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/293,150  
FILING DATE: 19-AUG-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 2-306745  
FILING DATE: 13-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Meuth, Donna M  
REGISTRATION NUMBER: 36,607  
REFERENCE/DOCKET NUMBER: 029650-049  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-293-150-23

Query Match 9.5%; Score 4; DB 6; Length 2;  
Best Local Similarity 100.0%; Pred. No. 3.4e+06;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 18 A 18  
DB 1 A 1

## RESULT 11

US-08-476-075-1  
SEQUENCE 1, Application US/08476075  
GENERAL INFORMATION:  
APPLICANT: Titmas, Richard C.  
APPLICANT: Hansen, David E.  
APPLICANT: Hong, Wonyo  
APPLICANT: Booth, Paul M.  
APPLICANT: Powell, Michael J.  
APPLICANT: Rees, Anthony R.  
APPLICANT: Massey, Richard J.  
TITLE OF INVENTION: PEPTIDE ANALOGS AND THEIR USE AS HAPTENS  
TITLE OF INVENTION: TO ELICIT CATALYTIC ANTIBODIES  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue, 25th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/476,075  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Evans, Barry  
REGISTRATION NUMBER: 22,802  
REFERENCE/DOCKET NUMBER: 370068-2701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: C-terminal  
US-08-476-075-1

Query Match 9.5%; Score 4; DB 8; Length 2;  
Best Local Similarity 100.0%; Pred. No. 3.4e+06;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

18 A 18

QY 1 A 1

## RESULT 12

US-08-564-222-1  
SEQUENCE 1, Application US/08564222  
GENERAL INFORMATION:  
APPLICANT: RENIMEL, Isabelle  
APPLICANT: ANDRE, Patrice  
APPLICANT: REDZINIAK, Gerard  
TITLE OF INVENTION: UTILIZATION OF A PEPTIDE HAVING A LYSINE  
TITLE OF INVENTION: GROUP AND AN ALANINE GROUP IN THE TERMINAL POSITION FOR  
TITLE OF INVENTION: THE PREPARATION OF A DEPIGMENTING COMPOSITION, AND

```
; TITLE OF INVENTION: DEPIGMENTING COMPOSITION THUS OBTAINED
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LOWE PRICE LeBLANC & BECKER
; STREET: 99 Canal Center Plaza, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/564,222
; FILING DATE: 15-DEC-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Stephen A.
; REGISTRATION NUMBER: 26,527
; REFERENCE/DOCKET NUMBER: 574-040
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-684-1111
; TELEFAX: 703-684-1124
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-564-222-1

Query Match          9.5%; Score 4; DB 9; Length 2;
Best Local Similarity 100.0%; Pred. No. 3.4e+06;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
Db 2 A 2

RESULT 13
US-09-077-574A-23
; Sequence 23, Application US/09077574A
; GENERAL INFORMATION:
; APPLICANT: Panaccio, Michael
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC COMPOSITIONS
; FILE REFERENCE: DAVIE60.001APC
; CURRENT APPLICATION NUMBER: US/09/077,574A
; CURRENT FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 2
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
; US-09-077-574A-23

Query Match          9.5%; Score 4; DB 14; Length 2;
Best Local Similarity 100.0%; Pred. No. 3.4e+06;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
Db 2 A 2

RESULT 14
US-09-302-905-13
; Sequence 13, Application US/09302905
```

```
; GENERAL INFORMATION:
; APPLICANT: Li, Xianqiang
; TITLE OF INVENTION: Truncated Green Fluorescent Protein
; FILE REFERENCE: D6103
; CURRENT APPLICATION NUMBER: US/09/302,905
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: US 60/083,967
; EARLIER FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 23
; SEQ ID NO 13
; LENGTH: 2
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; LOCATION: 227..228
; OTHER INFORMATION: Amino acid sequence of C-terminus of truncated EGFP
; OTHER INFORMATION: deletion mutant Del228
; US-09-302-905-13

Query Match          9.5%; Score 4; DB 17; Length 2;
Best Local Similarity 100.0%; Pred. No. 3.4e+06;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
Db 1 A 1

RESULT 15
US-09-302-905A-13
; Sequence 13, Application US/09302905A
; GENERAL INFORMATION:
; APPLICANT: Li, Xianqiang
; TITLE OF INVENTION: Truncated Green Fluorescent Protein
; FILE REFERENCE: clon-076
; CURRENT APPLICATION NUMBER: US/09/302,905A
; CURRENT FILING DATE: 1999-04-03
; PRIOR APPLICATION NUMBER: 60/083,967
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 2
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fragment
; US-09-302-905A-13

Query Match          9.5%; Score 4; DB 17; Length 2;
Best Local Similarity 100.0%; Pred. No. 3.4e+06;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
Db 1 A 1

Search completed: January 8, 2003, 11:12:08
JOB time : 217.517 secs
```





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2003, 10:58:59 ; Search time 11.0948 Seconds  
(without alignments)  
249.131 Million cell updates/sec

Title: US-09-003-869-5  
Perfect score: 42  
Sequence: 1 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 39

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262194 seqs, 70879160 residues  
1 number of hits satisfying chosen parameters: 262194

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Pending Patents AA.New.\*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 4     | 9.5         | 3      | 1     | PCT-US02-21135-25  |
| 2          | 4     | 9.5         | 3      | 1     | PCT-US02-21135-27  |
| 3          | 4     | 9.5         | 3      | 1     | PCT-US02-21135-28  |
| 4          | 4     | 9.5         | 3      | 1     | PCT-US02-21135-34  |
| 5          | 4     | 9.5         | 3      | 1     | PCT-US02-21135-42  |
| 6          | 4     | 9.5         | 3      | 1     | PCT-US02-21135-45  |
| 7          | 4     | 9.5         | 3      | 1     | PCT-US02-21135-46  |
| 8          | 4     | 9.5         | 3      | 1     | PCT-US02-21135-73  |
| 9          | 4     | 9.5         | 3      | 1     | PCT-US02-21135-74  |
| 10         | 4     | 9.5         | 3      | 1     | PCT-US02-21135-83  |
| 11         | 4     | 9.5         | 3      | 1     | PCT-US02-21135-103 |
| 12         | 4     | 9.5         | 3      | 1     | PCT-US02-21135-115 |
| 13         | 4     | 9.5         | 3      | 1     | PCT-US02-21135-116 |
| 14         | 4     | 9.5         | 3      | 1     | PCT-US02-34987-30  |
| 15         | 4     | 9.5         | 3      | 1     | PCT-US02-24469-14  |
| 16         | 4     | 9.5         | 3      | 1     | PCT-US02-24469-18  |
| 17         | 4     | 9.5         | 3      | 1     | PCT-US02-24469A-14 |
| 18         | 4     | 9.5         | 3      | 1     | PCT-US02-24469A-18 |
| 19         | 4     | 9.5         | 3      | 5     | US-09-580-893C-1   |
| 20         | 4     | 9.5         | 3      | 5     | US-09-580-893C-13  |
| 21         | 4     | 9.5         | 3      | 5     | US-09-580-893D-1   |
| 22         | 4     | 9.5         | 3      | 5     | US-09-580-893D-13  |
| 23         | 4     | 9.5         | 3      | 5     | US-09-853-080A-40  |
| 24         | 4     | 9.5         | 3      | 5     | US-09-653-812B-140 |
| 25         | 4     | 9.5         | 3      | 5     | US-09-653-812B-147 |
| 26         | 4     | 9.5         | 3      | 5     | US-09-653-812B-151 |

|    |   |     |   |   |                   |                   |
|----|---|-----|---|---|-------------------|-------------------|
| 27 | 4 | 9.5 | 3 | 6 | US-10-141-531-20  | Sequence 20, Appl |
| 28 | 4 | 9.5 | 3 | 6 | US-10-189-437-25  | Sequence 25, Appl |
| 29 | 4 | 9.5 | 3 | 6 | US-10-280-915-13  | Sequence 13, Appl |
| 30 | 4 | 9.5 | 3 | 6 | US-10-133-133A-9  | Sequence 9, Appl  |
| 31 | 4 | 9.5 | 3 | 6 | US-10-299-619-3   | Sequence 3, Appl  |
| 32 | 4 | 9.5 | 4 | 1 | PCT-US02-21135-1  | Sequence 1, Appl  |
| 33 | 4 | 9.5 | 4 | 1 | PCT-US02-21135-2  | Sequence 2, Appl  |
| 34 | 4 | 9.5 | 4 | 1 | PCT-US02-21135-3  | Sequence 3, Appl  |
| 35 | 4 | 9.5 | 4 | 1 | PCT-US02-21135-4  | Sequence 4, Appl  |
| 36 | 4 | 9.5 | 4 | 1 | PCT-US02-21135-5  | Sequence 5, Appl  |
| 37 | 4 | 9.5 | 4 | 1 | PCT-US02-21135-6  | Sequence 6, Appl  |
| 38 | 4 | 9.5 | 4 | 1 | PCT-US02-21135-7  | Sequence 7, Appl  |
| 39 | 4 | 9.5 | 4 | 1 | PCT-US02-21135-24 | Sequence 24, Appl |
| 40 | 4 | 9.5 | 4 | 1 | PCT-US02-21135-37 | Sequence 37, Appl |
| 41 | 4 | 9.5 | 4 | 1 | PCT-US02-21135-38 | Sequence 38, Appl |
| 42 | 4 | 9.5 | 4 | 1 | PCT-US02-21135-39 | Sequence 39, Appl |
| 43 | 4 | 9.5 | 4 | 1 | PCT-US02-21135-40 | Sequence 40, Appl |
| 44 | 4 | 9.5 | 4 | 1 | PCT-US02-21135-41 | Sequence 41, Appl |
| 45 | 4 | 9.5 | 4 | 1 | PCT-US02-21135-41 | Sequence 41, Appl |

## ALIGNMENTS

RESULT 1  
PCT-US02-21135-25  
; Sequence 25, Application PC/TUS0221135  
; GENERAL INFORMATION:  
; APPLICANT: MEDAREX, INC.  
; APPLICANT: Bebbington, Christopher R.  
; APPLICANT: Neider, Matthew H.  
; APPLICANT: Cardarelli, Pina M.  
; APPLICANT: Gangwar, Sanjeev  
; APPLICANT: Pickford, Lesley B.  
; APPLICANT: Pan, Chin  
; TITLE OF INVENTION: CD10-Activated Prodrug Compounds  
; FILE REFERENCE: COUL-023/01WO  
; CURRENT APPLICATION NUMBER: PCT/US02/21135  
; PRIOR FILING DATE: 2002-06-11  
; PRIOR APPLICATION NUMBER: US 60/297,596  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 124  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 25  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
PCT-US02-21135-25  
  
Query Match 9.5%; Score 4; DB 1; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 18 A 18  
Db 2 A 2  
  
RESULT 2  
PCT-US02-21135-27  
; Sequence 27, Application PC/TUS0221135  
; GENERAL INFORMATION:  
; APPLICANT: MEDAREX, INC.  
; APPLICANT: Bebbington, Christopher R.  
; APPLICANT: Neider, Matthew H.  
; APPLICANT: Cardarelli, Pina M.  
; APPLICANT: Gangwar, Sanjeev  
; APPLICANT: Pickford, Lesley B.  
; APPLICANT: Pan, Chin  
; TITLE OF INVENTION: CD10-Activated Prodrug Compounds  
; FILE REFERENCE: COUL-023/01WO

; CURRENT APPLICATION NUMBER: PCT/US02/21135  
; PRIOR FILING DATE: 2002-06-11  
; PRIOR APPLICATION NUMBER: US 60/297,596  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 124  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 27  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
PCT-US02-21135-27

Query Match 9.5%; Score 4; DB 1; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18  
|  
Db 2 A 2

RESULT 3  
PCT-US02-21135-28  
; Sequence 28, Application PC/TUS0221135  
; GENERAL INFORMATION:  
; APPLICANT: MEDAREX, INC.  
; APPLICANT: Bebbington, Christopher R.  
; APPLICANT: Neider, Matthew H.  
; APPLICANT: Cardarelli, Pina M.  
; APPLICANT: Gangwar, Sanjeev  
; APPLICANT: Pickford, Lesley B.  
; APPLICANT: Pan, Chin  
; TITLE OF INVENTION: CD10-Activated Prodrug Compounds  
; FILE REFERENCE: COUL-023/01WO  
; CURRENT APPLICATION NUMBER: PCT/US02/21135  
; CURRENT FILING DATE: 2002-06-11  
; PRIOR APPLICATION NUMBER: US 60/297,596  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 124  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 28  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
US02-21135-28

Query Match 9.5%; Score 4; DB 1; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18  
|  
Db 2 A 2

RESULT 4  
PCT-US02-21135-34  
; Sequence 34, Application PC/TUS0221135  
; GENERAL INFORMATION:  
; APPLICANT: MEDAREX, INC.  
; APPLICANT: Bebbington, Christopher R.  
; APPLICANT: Neider, Matthew H.  
; APPLICANT: Cardarelli, Pina M.  
; APPLICANT: Gangwar, Sanjeev  
; APPLICANT: Pickford, Lesley B.  
; APPLICANT: Pan, Chin  
; TITLE OF INVENTION: CD10-Activated Prodrug Compounds  
; FILE REFERENCE: COUL-023/01WO  
; CURRENT APPLICATION NUMBER: PCT/US02/21135

; CURRENT FILING DATE: 2002-06-11  
; PRIOR APPLICATION NUMBER: US 60/297,596  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 124  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 34  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
PCT-US02-21135-34

Query Match 9.5%; Score 4; DB 1; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18  
|  
Db 2 A 2

RESULT 5  
PCT-US02-21135-42  
; Sequence 42, Application PC/TUS0221135  
; GENERAL INFORMATION:  
; APPLICANT: MEDAREX, INC.  
; APPLICANT: Bebbington, Christopher R.  
; APPLICANT: Neider, Matthew H.  
; APPLICANT: Cardarelli, Pina M.  
; APPLICANT: Gangwar, Sanjeev  
; APPLICANT: Pickford, Lesley B.  
; APPLICANT: Pan, Chin  
; TITLE OF INVENTION: CD10-Activated Prodrug Compounds  
; FILE REFERENCE: COUL-023/01WO  
; CURRENT APPLICATION NUMBER: PCT/US02/21135  
; CURRENT FILING DATE: 2002-06-11  
; PRIOR APPLICATION NUMBER: US 60/297,596  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 124  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 42  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (1)..(1)  
; OTHER INFORMATION: Succinyl-Isoleucine  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (3)..(3)  
; OTHER INFORMATION: Leucine-doxorubicin  
PCT-US02-21135-42

Query Match 9.5%; Score 4; DB 1; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18  
|  
Db 2 A 2

RESULT 6  
PCT-US02-21135-45  
; Sequence 45, Application PC/TUS0221135  
; GENERAL INFORMATION:  
; APPLICANT: MEDAREX, INC.  
; APPLICANT: Bebbington, Christopher R.  
; APPLICANT: Neider, Matthew H.

; APPLICANT: Cardarelli, Pina M.  
; APPLICANT: Gangwar, Sanjeev  
; APPLICANT: Pickford, Lesley B.  
; APPLICANT: Pan, Chin  
; TITLE OF INVENTION: CD10-Activated Prodrug Compounds  
; FILE REFERENCE: COUL-023/01WO  
; CURRENT APPLICATION NUMBER: PCT/US02/21135  
; CURRENT FILING DATE: 2002-06-11  
; PRIOR APPLICATION NUMBER: US 60/297,596  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 124  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 45  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (1)..(1)  
; OTHER INFORMATION: Succinyl-Leucine  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (3)..(3)  
; OTHER INFORMATION: Leucine-doxorubicin  
PCT-US02-21135-45

Query Match 9.5%; Score 4; DB 1; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
|  
Db 2 A 2

RESULT 7  
PCT-US02-21135-46  
; Sequence 46, Application PC/TUS0221135  
; GENERAL INFORMATION:  
; APPLICANT: MEDAREX, INC.  
; APPLICANT: Bebbington, Christopher R.  
; APPLICANT: Neider, Matthew H.  
; APPLICANT: Cardarelli, Pina M.  
; APPLICANT: Gangwar, Sanjeev  
; APPLICANT: Pickford, Lesley B.  
; APPLICANT: Pan, Chin  
; TITLE OF INVENTION: CD10-Activated Prodrug Compounds  
; FILE REFERENCE: COUL-023/01WO  
; CURRENT APPLICATION NUMBER: PCT/US02/21135  
; CURRENT FILING DATE: 2002-06-11  
; PRIOR APPLICATION NUMBER: US 60/297,596  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 124  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 46  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (1)..(1)  
; OTHER INFORMATION: Succinyl-Leucine  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (3)..(3)  
; OTHER INFORMATION: Glycine-doxorubicin  
PCT-US02-21135-46

Query Match 9.5%; Score 4; DB 1; Length 3;

Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 18 A 18  
|  
Db 2 A 2  
  
RESULT 8  
PCT-US02-21135-73  
; Sequence 73, Application PC/TUS0221135  
; GENERAL INFORMATION:  
; APPLICANT: MEDAREX, INC.  
; APPLICANT: Bebbington, Christopher R.  
; APPLICANT: Neider, Matthew H.  
; APPLICANT: Cardarelli, Pina M.  
; APPLICANT: Gangwar, Sanjeev  
; APPLICANT: Pickford, Lesley B.  
; APPLICANT: Pan, Chin  
; TITLE OF INVENTION: CD10-Activated Prodrug Compounds  
; FILE REFERENCE: COUL-023/01WO  
; CURRENT APPLICATION NUMBER: PCT/US02/21135  
; CURRENT FILING DATE: 2002-06-11  
; PRIOR APPLICATION NUMBER: US 60/297,596  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 124  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 73  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (1)..(1)  
; OTHER INFORMATION: Succinyl-Methionine  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (3)..(3)  
; OTHER INFORMATION: Leucine-doxorubicin  
PCT-US02-21135-73

Query Match 9.5%; Score 4; DB 1; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
|  
Db 2 A 2

RESULT 9  
PCT-US02-21135-74  
; Sequence 74, Application PC/TUS0221135  
; GENERAL INFORMATION:  
; APPLICANT: MEDAREX, INC.  
; APPLICANT: Bebbington, Christopher R.  
; APPLICANT: Neider, Matthew H.  
; APPLICANT: Cardarelli, Pina M.  
; APPLICANT: Gangwar, Sanjeev  
; APPLICANT: Pickford, Lesley B.  
; APPLICANT: Pan, Chin  
; TITLE OF INVENTION: CD10-Activated Prodrug Compounds  
; FILE REFERENCE: COUL-023/01WO  
; CURRENT APPLICATION NUMBER: PCT/US02/21135  
; CURRENT FILING DATE: 2002-06-11  
; PRIOR APPLICATION NUMBER: US 60/297,596  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 124  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 74  
; LENGTH: 3

; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (1)..(1)  
; OTHER INFORMATION: Succinyl-Phenylalanine  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (3)..(3)  
; OTHER INFORMATION: Leucine-doxorubicin  
PCT-US02-21135-74

Query Match 9.5%; Score 4; DB 1; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18  
|  
2 A 2

RESULT 10  
PCT-US02-21135-83  
; Sequence 83, Application PC/TUS0221135  
; GENERAL INFORMATION:  
; APPLICANT: MEDAREX, INC.  
; APPLICANT: Bebbington, Christopher R.  
; APPLICANT: Neider, Matthew H.  
; APPLICANT: Cardarelli, Pina M.  
; APPLICANT: Gangwar, Sanjeev  
; APPLICANT: Pickford, Lesley B.  
; APPLICANT: Pan, Chin  
; TITLE OF INVENTION: CD10-Activated Prodrug Compounds  
; FILE REFERENCE: COUL-023/01WO  
; CURRENT APPLICATION NUMBER: PCT/US02/21135  
; CURRENT FILING DATE: 2002-06-11  
; PRIOR APPLICATION NUMBER: US 60/297,596  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 124  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 83  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (1)..(1)  
; OTHER INFORMATION: Succinyl-4-(Aminomethyl)benzoic acid  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (4)..(4)  
; OTHER INFORMATION: Leucine-daunorubicin  
PCT-US02-21135-83

Query Match 9.5%; Score 4; DB 1; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18  
|  
3 A 3

RESULT 11  
PCT-US02-21135-103  
; Sequence 103, Application PC/TUS0221135  
; GENERAL INFORMATION:  
; APPLICANT: MEDAREX, INC.  
; APPLICANT: Bebbington, Christopher R.

; APPLICANT: Neider, Matthew H.  
; APPLICANT: Cardarelli, Pina M.  
; APPLICANT: Gangwar, Sanjeev  
; APPLICANT: Pickford, Lesley B.  
; APPLICANT: Pan, Chin  
; TITLE OF INVENTION: CD10-Activated Prodrug Compounds  
; FILE REFERENCE: COUL-023/01WO  
; CURRENT APPLICATION NUMBER: PCT/US02/21135  
; CURRENT FILING DATE: 2002-06-11  
; PRIOR APPLICATION NUMBER: US 60/297,596  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 124  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 103  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (1)..(1)  
; OTHER INFORMATION: Succinyl-Isoleucine  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (3)..(3)  
; OTHER INFORMATION: Leucine-doxorubicin  
PCT-US02-21135-103

Query Match 9.5%; Score 4; DB 1; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18  
|  
2 A 2

RESULT 12  
PCT-US02-21135-115  
; Sequence 115, Application PC/TUS0221135  
; GENERAL INFORMATION:  
; APPLICANT: MEDAREX, INC.  
; APPLICANT: Bebbington, Christopher R.  
; APPLICANT: Neider, Matthew H.  
; APPLICANT: Cardarelli, Pina M.  
; APPLICANT: Gangwar, Sanjeev  
; APPLICANT: Pickford, Lesley B.  
; APPLICANT: Pan, Chin  
; TITLE OF INVENTION: CD10-Activated Prodrug Compounds  
; FILE REFERENCE: COUL-023/01WO  
; CURRENT APPLICATION NUMBER: PCT/US02/21135  
; CURRENT FILING DATE: 2002-06-11  
; PRIOR APPLICATION NUMBER: US 60/297,596  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 124  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 115  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (1)..(1)  
; OTHER INFORMATION: 9-Fluorenylmethylloxycarbonyl-Isoleucine  
PCT-US02-21135-115

Query Match 9.5%; Score 4; DB 1; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
|  
Db 2 A 2

## RESULT 13

PCT-US02-21135-116  
; Sequence 116, Application PC/TUS0221135  
; GENERAL INFORMATION:  
; APPLICANT: MEDAREX, INC.  
; APPLICANT: Bebbington, Christopher R.  
; APPLICANT: Neider, Matthew H.  
; APPLICANT: Cardarelli, Pina M.  
; APPLICANT: Gangwar, Sanjeev  
; APPLICANT: Pickford, Lesley B.  
; APPLICANT: Pan, Chin  
; TITLE OF INVENTION: CD10-Activated Prodrug Compounds  
; FILE REFERENCE: COUL-023/01WO  
; CURRENT APPLICATION NUMBER: PCT/US02/21135  
; PRIOR FILING DATE: 2002-06-11  
; PRIOR APPLICATION NUMBER: US 60/297,596  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 124  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 116  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (1)..(1)  
; OTHER INFORMATION: 9-Fluorenylmethoxycarbonyl-Isoleucine  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (3)..(3)  
; OTHER INFORMATION: Leucine-doxorubicin  
PCT-US02-21135-116

Query Match 9.5%; Score 4; DB 1; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
|  
Db 2 A 2

## RESULT 14

PCT-US02-34987-30  
; Sequence 30, Application PC/TUS0234987  
; GENERAL INFORMATION:  
; APPLICANT: Board of Regents, The University of Texas System (applicant for the  
; APPLICANT: purposes of all designated states except US)  
; APPLICANT: Arap, Wadhi (applicant for the purpose of the United States of America  
; APPLICANT: only)  
; APPLICANT: Kolonin, Mikhail G. (applicant for the purpose of the United States of  
; APPLICANT: America only)  
; APPLICANT: Mintz, Paul J. (applicant for the purpose of the United States of America  
; APPLICANT: only)  
; APPLICANT: Pasqualini, Renata (applicant for the purpose of the United States of  
; APPLICANT: America only)  
; APPLICANT: Zurita, Amado J. (applicant for the purpose of the United States of America  
; APPLICANT: only)  
; TITLE OF INVENTION: Compositions and Methods of Use of Targeting Peptides for Diagnos  
; FILE REFERENCE: 005774.P010PCT  
; CURRENT APPLICATION NUMBER: PCT/US02/34987  
; PRIOR FILING DATE: 2002-10-30  
; PRIOR APPLICATION NUMBER: PCT/US02/27836  
; PRIOR FILING DATE: 2001-08-30  
; NUMBER OF SEQ ID NOS: 132

; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 30  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: Synthetic Peptide  
PCT-US02-34987-30

Query Match 9.5%; Score 4; DB 1; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
|  
Db 1 A 1

RESULT 15  
PCT-US02-24469-14  
; Sequence 14, Application PC/TUS0224469  
; GENERAL INFORMATION:  
; APPLICANT: Merck & Co., Inc.  
; TITLE OF INVENTION: Biological Assay Detection Method  
; FILE REFERENCE: PCT 20900Y  
; CURRENT APPLICATION NUMBER: PCT/US02/24469  
; CURRENT FILING DATE: 2002-08-02  
; PRIOR APPLICATION NUMBER: 60/310,599  
; PRIOR FILING DATE: 2001-08-07  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 3  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic peptide  
PCT-US02-24469-14

Query Match 9.5%; Score 4; DB 1; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
|  
Db 3 A 3

Search completed: January 8, 2003, 11:12:47  
Job time : 11.0948 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2003, 10:54:49 ; Search time 12.4397 Seconds  
(without alignments)  
301.394 Million cell updates/sec

Title: US-09-003-869-5

Perfect score: 42

Sequence: 1 XX 39

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID  | Description        |
|------------|-------|-------------|--------|--------|--------------------|
| 1          | 4     | 9.5         | 3      | PT0636 | T-cell receptor be |
| 2          | 4     | 9.5         | 3      | PT0578 | T-cell receptor be |
| 3          | 4     | 9.5         | 3      | PT0571 | T-cell receptor be |
| 4          | 4     | 9.5         | 3      | PT0622 | T-cell receptor be |
| 5          | 4     | 9.5         | 3      | S68328 | blood cell protein |
| 6          | 4     | 9.5         | 4      | S18401 | thyroglobulin - do |
| 7          | 4     | 9.5         | 4      | PL0146 | carbon-monoxide de |
| 8          | 4     | 9.5         | 4      | A48360 | gamma subunit of p |
| 9          | 4     | 9.5         | 4      | S43014 | hypothetical prote |
| 10         | 4     | 9.5         | 4      | A27897 | glucan 1,4-alpha-g |
| 11         | 4     | 9.5         | 4      | JQ1273 | neuropeptide Antho |
| 12         | 4     | 9.5         | 4      | A34626 | RPCH-related neuro |
| 13         | 4     | 9.5         | 4      | A32480 | achatin-I - giant  |
| 14         | 4     | 9.5         | 4      | S39390 | myosin-light-chain |
| 15         | 4     | 9.5         | 4      | I61883 | protamine P1 - ora |
| 16         | 4     | 9.5         | 4      | PT0271 | Ig heavy chain CRD |
| 17         | 4     | 9.5         | 4      | S43959 | Ig mu chain V regi |
| 18         | 4     | 9.5         | 4      | PT0696 | T-cell receptor be |
| 19         | 4     | 9.5         | 4      | PT0645 | T-cell receptor be |
| 20         | 4     | 9.5         | 4      | PT0633 | T-cell receptor be |
| 21         | 4     | 9.5         | 4      | PT0711 | T-cell receptor be |
| 22         | 4     | 9.5         | 4      | PT0661 | T-cell receptor be |
| 23         | 4     | 9.5         | 4      | PT0712 | T-cell receptor be |
| 24         | 4     | 9.5         | 4      | PT0534 | T-cell receptor be |
| 25         | 4     | 9.5         | 4      | PT0698 | T-cell receptor be |
| 26         | 4     | 9.5         | 4      | PT0677 | T-cell receptor be |
| 27         | 4     | 9.5         | 4      | PT0551 | T-cell receptor be |
| 28         | 4     | 9.5         | 4      | PT0697 | T-cell receptor be |
| 29         | 4     | 9.5         | 4      | PT0706 | T-cell receptor be |

|    |   |     |   |   |        |                     |
|----|---|-----|---|---|--------|---------------------|
| 30 | 4 | 9.5 | 4 | 2 | PT0675 | T-cell receptor be  |
| 31 | 4 | 9.5 | 4 | 2 | PT0721 | T-cell receptor be  |
| 32 | 4 | 9.5 | 4 | 2 | PT0566 | T-cell receptor be  |
| 33 | 4 | 9.5 | 4 | 2 | A26209 | protein-glutamine   |
| 34 | 4 | 9.5 | 4 | 2 | I37013 | protamine P1 - Cer  |
| 35 | 4 | 9.5 | 4 | 2 | I84439 | protamine P1 - sav  |
| 36 | 4 | 9.5 | 5 | 2 | JN0860 | peptidyl-dipeptida  |
| 37 | 4 | 9.5 | 5 | 2 | C23751 | spinal cord peptid  |
| 38 | 4 | 9.5 | 5 | 2 | A41225 | copper resistance   |
| 39 | 4 | 9.5 | 5 | 2 | B41225 | copper resistance   |
| 40 | 4 | 9.5 | 5 | 2 | S70154 | URF2 protein - Xan  |
| 41 | 4 | 9.5 | 5 | 2 | I40702 | primase - Citrobac  |
| 42 | 4 | 9.5 | 5 | 2 | I40698 | biotin B - Citrobac |
| 43 | 4 | 9.5 | 5 | 2 | I39964 | ribosomal protein   |
| 44 | 4 | 9.5 | 5 | 2 | I39966 | ribosomal protein   |
| 45 | 4 | 9.5 | 5 | 2 | I39965 | ribosomal protein   |

ALIGNMENTS

RESULT 1

PT0636

T-cell receptor beta chain V-D-J region (100-2AT) - mouse (fragment)

C:Species: Mus musculus (house mouse)

C>Date: 15-Jun-2001 #sequence\_revision 15-Jun-2001 #text\_change 15-Jun-2001

C:Accession: PT0636

R:Feeney, A.J.

J. Exp. Med. 174, 115-124, 1991

A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions

A:Reference number: PT0509; MUID:91277601; PMID:1711558

A:Accession: PT0636

A:Molecule type: mRNA

A:Residues: 1-3 <FEE>

A:Experimental source: newborn thymus, strain BALB/c

C:Keywords: T-cell receptor

Query Match 9.5%; Score 4; DB 3; Length 3;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18

Db 1 A 1

RESULT 2

PT0578

T-cell receptor beta chain V-D-J region (141-1BD) - mouse (fragment)

C:Species: Mus musculus (house mouse)

C>Date: 15-Jun-2001 #sequence\_revision 15-Jun-2001 #text\_change 15-Jun-2001

C:Accession: PT0578

R:Feeney, A.J.

J. Exp. Med. 174, 115-124, 1991

A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions

A:Reference number: PT0509; MUID:91277601; PMID:1711558

A:Accession: PT0578

A:Status: translation not shown

A:Molecule type: mRNA

A:Residues: 1-3 <FEE>

A:Experimental source: day 19 fetal thymus, strain BALB/c

C:Keywords: T-cell receptor

Query Match 9.5%; Score 4; DB 3; Length 3;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18

Db 1 A 1

RESULT 3  
PT0571  
T-cell receptor beta chain V-D-J region (141-ICM) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 15-Jun-2001 #sequence\_revision 15-Jun-2001 #text\_change 15-Jun-2001  
C:Accession: PT0571  
R:Feeney, A.J.  
J. Exp. Med. 174, 115-124, 1991  
A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
A:Reference number: PT0509; MUID:91277601; PMID:1711558  
A:Accession: PT0571  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-3 <FEE>  
A:Experimental source: day 19 fetal thymus, strain BALB/c  
C:Keywords: T-cell receptor

Query Match 9.5%; Score 4; DB 3; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
|  
Db 1 A 1

RESULT 4  
PT0622  
T-cell receptor beta chain V-D-J region - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 15-Jun-2001 #sequence\_revision 15-Jun-2001 #text\_change 15-Jun-2001  
C:Accession: PT0622; PT0680; PT0582; PT0673  
R:Feeney, A.J.  
J. Exp. Med. 174, 115-124, 1991  
A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
A:Reference number: PT0509; MUID:91277601; PMID:1711558  
A:Accession: PT0622  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-3 <FEE>  
A:Experimental source: newborn thymus, strain BALB/c, clone 111-1p  
A:Accession: PT0680  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-3 <FEE1>  
A:Experimental source: day 18 fetal thymus, strain BALB/c clones 154-1J and 154-1P  
A:Accession: PT0582  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-3 <FEE2>  
A:Experimental source: day 19 fetal thymus, strain BALB/c, clone 159-2A  
C:Keywords: T-cell receptor

Query Match 9.5%; Score 4; DB 3; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
|  
Db 1 A 1

RESULT 5  
S68328  
blood cell protein A - Molgula manhattensis (fragment)  
C:Species: Molgula manhattensis  
C:Date: 15-Jun-2001 #sequence\_revision 15-Jun-2001 #text\_change 15-Jun-2001  
C:Accession: S68328  
R:Taylor, S.W.; Ross, M.M.; Waite, J.H.  
Arch. Biochem. Biophys. 324, 228-240, 1995  
A:Title: Novel 3,4-di- and 3,4,5-trihydroxyphenylalanine-containing polypeptides from the  
A:Reference number: S68325; MUID:96132650; PMID:8554314  
A:Accession: S68328

Query Match 9.5%; Score 4; DB 3; Length 3;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
|  
Db 1 A 1

RESULT 6  
S18401  
thyroglobulin - dog (fragment)  
C:Species: Canis lupus familiaris (dog)  
C:Date: 19-Mar-1997 #sequence\_revision 18-Aug-2000 #text\_change 18-Aug-2000  
C:Accession: S18401  
R:Donda, A.; Vassart, G.; Christophe, D.  
Biochim. Biophys. Acta 1090, 235-237, 1991  
A:Title: Isolation and characterization of the canine thyroglobulin gene promoter region  
A:Reference number: S18401; MUID:92031697; PMID:1932116  
A:Accession: S18401  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-4 <DON>  
A:Cross-references: GB:S61184; NID:g237714; PIDN:AAB20127.1; PID:g237715  
C:Superfamily: thyroglobulin; cholinesterase homology; thyroglobulin type I repeat homology  
C:Keywords: duplication; iodine; thyroid gland; thyroid hormone biosynthesis

Query Match 9.5%; Score 4; DB 2; Length 4;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
|  
Db 2 A 2

RESULT 7  
PL0146  
carbon-monoxide dehydrogenase (EC 1.2.99.2) small chain - Pseudomonas carboxydohydrog  
C:Species: Pseudomonas carboxydohydrog  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 28-Apr-1993  
C:Accession: PL0146  
R:Kraut, M.; Hugendieck, I.; Herwig, S.; Meyer, O.  
Arch. Microbiol. 152, 335-341, 1989  
A:Title: Homology and distribution of CO dehydrogenase structural genes in carboxydot  
A:Reference number: PL0138; MUID:90055678; PMID:2818128  
A:Accession: PL0146  
A:Molecule type: protein  
A:Residues: 1-4 <KRA>  
C:Comment: Carbon-monoxide dehydrogenase consists of three polypeptide chains: large,  
C:Keywords: oxidoreductase

Query Match 9.5%; Score 4; DB 2; Length 4;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
|  
Db 2 A 2

RESULT 8  
A48360  
gamma subunit of Protein A - Methylosinus trichosporium (fragment)  
C:Species: Methylosinus trichosporium  
C:Date: 19-Nov-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Oct-1999  
C:Accession: A48360  
R:Cardy, D.L.; Laidler, V.; Salmond, G.P.; Murrell, J.C.  
Arch. Microbiol. 156, 477-483, 1991  
A:Title: The methane monooxygenase gene cluster of Methylosinus trichosporium: clonin



A:Reference number: A48360; MUID:92153031; PMID:1785954

A:Contents: 0B3b

A:Accession: A48360

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-4 <CAR>

A:Cross-references: GB:S81887; NID:9245213; PIDN:AAB21391.1; PID:9245214

A>Note: sequence extracted from NCBI backbone (NCBIN:81887, NCBIP:81912)

Query Match 9.5%; Score 4; DB 2; Length 4;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18

Db 4 A 4

#### RESULT 9

3014

C:Species: Bacterial protein URF-2Y - Yersinia enterocolitica transposon TN3926

C:Accession: S43014

C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 08-Oct-1999

R:Osborn, S.E.V.; Turner, A.K.; Grinstead, J.

A:Description: The structure of the bacterial transposable element, Tn3926.

A:Reference number: S43011

A:Molecule type: DNA

A:Residues: 1-4 <OSB>

A:Cross-references: EMBL:X78059; NID:g460067; PIDN:CAA54980.1; PID:g581836

C:Genetics:

A:Mobile element: transposon TN3926

Query Match 9.5%; Score 4; DB 2; Length 4;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18

Db 3 A 3

#### RESULT 10

A27897 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - Aspergillus phoenicis (fragment)

Alternate names: glucanase

Species: Aspergillus phoenicis

Date: 16-Aug-1988 #sequence\_revision 16-Aug-1988 #text\_change 06-Dec-1996

C:Accession: A27897

R:Inokuchi, N.; Takahashi, T.; Irie, M.

J. Biochem. 90, 1055-1067, 1981

A>Title: Purification and characterization of a minor glucanase from Aspergillus saitoi

A:Reference number: A27897; MUID:82075730; PMID:6796572

A>Note: Aspergillus saitoi

A:Molecule type: protein

A:Residues: 1-4 <INO>

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 9.5%; Score 4; DB 2; Length 4;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18

Db 1 A 1

#### RESULT 11

JQ1273

neuropeptide Antho-KAamide - sea anemone (Anthopleura elegantissima)

C:Species: Anthopleura elegantissima

C>Date: 31-Mar-1992 #sequence\_revision 04-Dec-1992 #text\_change 08-Dec-1995

C:Accession: JQ1273

R:Notthacker, H.P.; Rinehart, K.L.; Grimmelikhuijzen, C.J.P.

Biochem. Biophys. Res. Commun. 179, 1205-1211, 1991

A>Title: Isolation of L-3-phenyllactyl-Phe-Lys-Ala-NH2 (Antho-KAamide), a novel neuro

A:Reference number: JQ1273; MUID:92028852; PMID:1681803

A:Accession: JQ1273

A:Molecule type: protein

A:Residues: 1-4 <NOT>

C:Comment: The carboxyl-terminal amide probably arises from cleavage of a following g

C:Keywords: amidated carboxyl end; neuropeptide; phenyllactylation

F:1/Modified site: L-3-phenyllactic acid (Phe) #status experimental

F:4/Modified site: amidated carboxyl end (Ala) #status experimental

Query Match 9.5%; Score 4; DB 2; Length 4;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18

Db 4 A 4

#### RESULT 12

A34626

RPCH-related neuropeptide - ferruginous spindle

C:Species: Fusinus ferrugineus (ferruginous spindle)

C>Date: 06-Jul-1990 #sequence\_revision 06-Jul-1990 #text\_change 31-Dec-1993

C:Accession: A34626

R:Kuroki, Y.; Kanda, T.; Kubota, I.; Fujisawa, Y.; Ikeda, T.; Miura, A.; Minamitake,

Biochem. Biophys. Res. Commun. 167, 273-279, 1990

A>Title: A molluscan neuropeptide related to the crustacean hormone, RPCH.

A:Reference number: A34626; MUID:90179762; PMID:2310394

A:Accession: A34626

A>Status: preliminary

A:Molecule type: protein

A:Residues: 1-4 <KUR>

C:Keywords: neuropeptide

Query Match 9.5%; Score 4; DB 2; Length 4;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18

Db 1 A 1

#### RESULT 13

A32480

achatin-I - giant African snail

N:Contains: achatin-II

C:Species: Achatina fulica (giant African snail)

C>Date: 12-Oct-1989 #sequence\_revision 12-Oct-1989 #text\_change 17-Mar-1999

C:Accession: A32480

R:Kamatani, Y.; Minakata, R.; Kenny, P.T.M.; Iwashita, T.; Watanabe, K.; Funase, K.;

Biochem. Biophys. Res. Commun. 160, 1015-1020, 1989

A>Title: Achatin-I, an endogenous neuroexcitatory tetrapeptide from Achatina fulica f

A:Reference number: A32480; MUID:89273551; PMID:2597281

A:Accession: A32480

A:Molecule type: protein

A:Residues: 1-4 <AM>

A>Note: stereochemistry of the active form confirmed by chemical synthesis

R:Ishida, T.; In, Y.; Inoue, M.; Yasuda-Kamatani, Y.; Minakata, H.; Iwashita, T.; Nom

FEBS Lett. 307, 253-256, 1992

A>Title: Effect of the D-Phe(2)

(H-Gly-Phe-Ala-Asp-OH).

A:Reference number: A44691; MUID:92354723; PMID:1644179

A:Contents: annotation: X-ray crystallography, 0.85 angstroms

A>Note: achatin-II has L-phenylalanine

C:Keywords: D-amino acid

F:2/Modified site: D-phenylalanine (Phe) #status experimental

Query Match 9.5%; Score 4; DB 2; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
 Db 3 A 3

## RESULT 14

S39390  
 myosin-light-chain kinase (EC 2.7.1.117), smooth muscle - turkey (fragment)  
 C:Species: Meleagris gallopavo (common turkey)  
 C>Date: 19-Mar-1997 #sequence\_revision 10-Oct-1997 #text\_change 07-May-1999  
 C:Accession: S39390  
 R:Komatsu, H.; Ikebe, M.  
 Biochem. J. 296, 53-58, 1993  
 A:Title: Affinity labelling of smooth-muscle myosin light-chain kinase with 5'-(p-(fluoromethyl)phosphoryl)-ATP  
 Reference number: S39390; MUID:94071841; PMID:8250857  
 Accession: S39390  
 R:Molecule type: protein  
 A:Residues: 1-4 <KOM>  
 A:Experimental source: gizzard  
 C:Keywords: phosphotransferase; smooth muscle

Query Match 9.5%; Score 4; DB 2; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
 Db 1 A 1

## RESULT 15

I61883  
 protamine P1 - orangutan. (fragment)  
 C:Species: Pongo pygmaeus (orangutan)  
 C>Date: 06-Sep-1996 #sequence\_revision 06-Sep-1996 #text\_change 21-Jul-2000  
 C:Accession: I61883  
 R:Queralt, R.; Oliva, R.  
 Gene 133, 197-204, 1993  
 A:Title: Identification of conserved potential regulatory sequences of the protamine-encoding gene P1 from Pongo pygmaeus  
 Reference number: I37013; MUID:94040810; PMID:8224908  
 Accession: I61883  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 R:Molecule type: DNA  
 A:Residues: 1-4 <RES>  
 Cross-references: EMBL:212146; NID:g38156; PIDN:CAA78130.1; PID:g4379372

Query Match 9.5%; Score 4; DB 2; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 A 18  
 Db 2 A 2

Search completed: January 8, 2003, 10:55:42  
 Job time : 14.4397 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2003, 10:54:49 ; Search time 6.72414 Seconds  
(without alignments)  
240.563 Million cell updates/sec

Title: US-09-003-869-5

Perfect score: 42

Sequence: 1 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 39

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

cal number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 4     | 9.5         | 4      | 1     | ACH1_ACHF   |
| 2          | 4     | 9.5         | 4      | 1     | DCM5_PSECH  |
| 3          | 4     | 9.5         | 4      | 1     | FFKA_ATEL   |
| 4          | 4     | 9.5         | 5      | 1     | BIOB_CITFR  |
| 5          | 4     | 9.5         | 5      | 1     | BP7_BOTIN   |
| 6          | 4     | 9.5         | 5      | 1     | RE11_LITRU  |
| 7          | 4     | 9.5         | 5      | 1     | RE21_LITRU  |
| 8          | 4     | 9.5         | 5      | 1     | SUGA_ACHDO  |
| 9          | 4     | 9.5         | 6      | 1     | CIP2_MYTED  |
| 10         | 4     | 9.5         | 6      | 1     | LOK1_LOCMI  |
| 11         | 4     | 9.5         | 6      | 1     | OYPL_LEPDE  |
| 12         | 4     | 9.5         | 6      | 1     | TRP1_PSEPU  |
| 13         | 4     | 9.5         | 6      | 1     | UN06_CLOPA  |
| 14         | 4     | 9.5         | 7      | 1     | ALL2_CARMA  |
| 15         | 4     | 9.5         | 7      | 1     | ALL3_CARMA  |
| 16         | 4     | 9.5         | 7      | 1     | ALL4_CARMA  |
| 17         | 4     | 9.5         | 7      | 1     | ALL5_CARMA  |
| 18         | 4     | 9.5         | 7      | 1     | BRHP_CONIM  |
| 19         | 4     | 9.5         | 7      | 1     | CARP_MYTED  |
| 20         | 4     | 9.5         | 7      | 1     | CIA_ENTFA   |
| 21         | 4     | 9.5         | 7      | 1     | EIO5_LITRU  |
| 22         | 4     | 9.5         | 7      | 1     | FAR3_HAECO  |
| 23         | 4     | 9.5         | 7      | 1     | FAR3_PANRE  |
| 24         | 4     | 9.5         | 7      | 1     | IGA0_DKDE   |
| 25         | 4     | 9.5         | 7      | 1     | MNP1_LEPDE  |
| 26         | 4     | 9.5         | 7      | 1     | UC24_MAIZE  |
| 27         | 4     | 9.5         | 7      | 1     | UF03_MOUSE  |
| 28         | 4     | 9.5         | 7      | 1     | UH11_RAT    |
| 29         | 4     | 9.5         | 8      | 1     | AL12_CARMA  |
| 30         | 4     | 9.5         | 8      | 1     | AL15_CARMA  |
| 31         | 4     | 9.5         | 8      | 1     | ALL4_CYDPO  |
| 32         | 4     | 9.5         | 8      | 1     | ALL5_CYDPO  |
| 33         | 4     | 9.5         | 8      | 1     | ALL7_CARMA  |

|    |   |     |   |   |            |
|----|---|-----|---|---|------------|
| 34 | 4 | 9.5 | 8 | 1 | ALL8_CARMA |
| 35 | 4 | 9.5 | 8 | 1 | ALL9_CARMA |
| 36 | 4 | 9.5 | 8 | 1 | B4K4_PORGI |
| 37 | 4 | 9.5 | 8 | 1 | CAD1_ENTFA |
| 38 | 4 | 9.5 | 8 | 1 | CLP_THICU  |
| 39 | 4 | 9.5 | 8 | 1 | FAR7_ASCSU |
| 40 | 4 | 9.5 | 8 | 1 | FAR8_CALVO |
| 41 | 4 | 9.5 | 8 | 1 | GLUR_HUMAN |
| 42 | 4 | 9.5 | 8 | 1 | LCK1_LEUMA |
| 43 | 4 | 9.5 | 8 | 1 | LCK4_LEUMA |
| 44 | 4 | 9.5 | 8 | 1 | LCK7_LEUMA |
| 45 | 4 | 9.5 | 8 | 1 | LCK8_LEUMA |

|        |             |
|--------|-------------|
| P81811 | carcinus ma |
| P81812 | carcinus ma |
| P81866 | porphyromon |
| P81268 | enterococcu |
| P80488 | thiobacilli |
| P43171 | ascaris suu |
| P41863 | calliphora  |
| P02729 | homo sapien |
| P21140 | leucophaea  |
| P21143 | leucophaea  |
| P19989 | leucophaea  |
| P19990 | leucophaea  |

#### ALIGNMENTS

RESULT 1  
ACH1\_ACHF  
ID ACH1\_ACHF STANDARD; PRT; 4 AA.  
AC P35904;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Achatin-I.  
OS Achatina fulica (Giant African snail).  
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;  
OC Achatinacea; Achatinidae; Achatina.  
OX NCBI\_TaxID=6530;  
RN [1]  
RP SEQUENCE CHARACTERIZATION, AND SYNTHESIS.  
RC STRAIN=FERUSSAC; TISSUE=Ganglion;  
RX MEDLINE=69273551; PubMed=2597281;  
RA Kamatani Y., Minakata H., Kenny P. T. M., Iwashita T., Watanabe K.,  
RA Funase K., Sun X. P., Yongsiri A., Kim K. H., Novales-Li P.,  
RA Novales E. T., Kanapi C. G., Takeuchi H., Nomoto K.;  
RT "Achatin-I, an endogenous neuroexcitatory tetrapeptide from Achatina  
RT fulica Ferussac containing a D-amino acid residue.";  
RL Biochem. Biophys. Res. Commun. 160:1015-1020(1989).  
RN [2]  
RP CHARACTERIZATION.  
RC STRAIN=FERUSSAC; TISSUE=Heart atrium;  
RX MEDLINE=91264856; PubMed=1675568;  
RA Fujimoto K., Kubota I., Yasuda-Kamatani Y., Minakata H., Nomoto K.,  
RA Yoshida M., Harada A., Muneoka Y., Kobayashi M.;  
RT "Purification of achatin-I from the atria of the African giant snail,  
RL Achatina fulica, and its possible function.";  
RL Biochem. Biophys. Res. Commun. 177:847-853(1991).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY.  
RX MEDLINE=93014529; PubMed=1399285;  
RA Iwashita T., In Y., Doi M., Inoue M., Yasuda-Kamatani Y., Minakata H.,  
RA Iwashita T., Nomoto K.;  
RT "Crystal structure and molecular conformation of achatin-I  
RT (H-Gly-D-Phe-Ala-Asp-OH), an endogenous neuropeptide containing a  
RT D-amino acid residue.";  
RL Int. J. Pept. Protein Res. 39:258-264(1992).  
CC -!- FUNCTION: NEUROEXCITATORY PEPTIDE; INCREASES THE IMPULSE FREQUENCY  
CC AND PRODUCES A SPIKE BROADENING OF THE IDENTIFIED HEART EXCITATORY  
CC NEURON (PON); ALSO ENHANCES THE AMPLITUDE AND FREQUENCY OF THE  
CC HEART BEAT. HAS ALSO AN EFFECT ON SEVERAL OTHER MUSCLES.  
DR PIR: A32480; A32480.  
KW Hormone; D-amino acid.  
FT MOD\_RES 2 D-PHENYLALANINE.  
SQ SEQUENCE 4 AA: 408 MW; 6AADD9C810000000 CRC64;

Query Match 9.5%; Score 4; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18

DB 3 A 3

```

RESULT 2
DCMS_PSECH STANDARD; PRT; 4 AA.
AC P19918;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Carbon monoxide dehydrogenase small chain (EC 1.2.99.2) (CO
DE dehydrogenase subunit S) (CO-DH S) (Fragment).
DE CUTS.
GN Pseudomonas carboxydohydrogena.
OS Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group.
OX NCBI_TaxID=290;
RN [1]
RP SEQUENCE
RX MEDLINE=90055678; PubMed=2818128;
RA Kraut M., Hugendieck I., Herwig S., Meyer O.;
RT "Homology and distribution of CO dehydrogenase structural genes in
RL Arch. Microbiol. 152:335-341(1989).
CC -!- FUNCTION: Catalyzes the oxidation of carbon monoxide to carbon
CC dioxide.
CC -!- CATALYTIC ACTIVITY: CO + H(2)O + acceptor = CO(2) + reduced
CC acceptor.
CC -!- COFACTOR: BINDS TWO 2FE-2S CLUSTERS.
CC -!- SUBUNIT: CONSISTS OF THREE POLYPEPTIDE CHAINS: LARGE, MEDIUM, AND
CC SMALL.
DR PIR, P01146; P01146.
KW Oxidoreductase; Iron-sulfur.
FT NON_TER 4
SQ SEQUENCE 4 AA; 420 MW; 6DD33DD6F000000000 CRC64;

Query Match 9.5%; Score 4; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
Db 2 A 2

RESULT 3
FFKA_ANTL STANDARD; PRT; 4 AA.
ID F58705;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Antho-Kaamide.
OS Anthopleura elegantissima (Sea anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nyantheae; Actiniidae; Anthopleura.
OX NCBI_TaxID=6110;
RN [1]
RP SEQUENCE
RX MEDLINE=92028852; PubMed=1681803;
RA Nethacker H.-P., Rinehart K.L. Jr.;
RT "Isolation of L-3-phenylactyl-Phe-Lys-Ala-NH2 (Antho-Kaamide), a
RL Biochem. Biophys. Res. Commun. 179:1205-1211(1991).
RN [2]
RP FUNCTION.
RX MEDLINE=93391436; PubMed=8397415;
RA McFarlane I.D., Hudnan D., Nethacker H.-P., Grimmelikhuijzen C.J.P.;
RT "The expansion behaviour of sea anemones may be coordinated by two
RT inhibitory neuropeptides, Antho-Kaamide and Antho-Riamide.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 253:183-188(1993).
CC -!- FUNCTION: Inhibits spontaneous contractions in several muscle
CC groups. May be involved in the expansion phase of feeding
CC behaviour in sea anemones.

```

```

CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Neuron-specific.
KW Neuropeptide; Amidation.
FT MOD_RES 1 L-3-PHENYLLACTYL.
FT MOD_RES 4 4 AMIDATION.
SQ SEQUENCE 4 AA; 512 MW; 6DD339C9A000000000 CRC64;

Query Match 9.5%; Score 4; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
Db 4 A 4

RESULT 4
BIOB_CITFR STANDARD; PRT; 5 AA.
ID AC P12997;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Biotin synthase (EC 2.8.1.6) (Biotin synthetase) (Fragment).
DE GN BIOB.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89006280; PubMed=2971595;
RA Shiuan D., Campbell A.;
RT "Transcriptional regulation and gene arrangement of Escherichia coli,
RT Citrobacter freundii and Salmonella typhimurium biotin operons.";
RL Gene 67:203-211(1988).
CC -!- CATALYTIC ACTIVITY: Dethiobiotin + sulfur = biotin.
CC -!- PATHWAY: Biotin biosynthesis; last step.
CC -!- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M21922; -; NOT_ANNOTATED_CDS.
CC Biotin biosynthesis; Iron-sulfur; Transferase.
KW NON_TER 5
SQ SEQUENCE 5 AA; 532 MW; 75A5B1EDD6F000000000 CRC64;

Query Match 9.5%; Score 4; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
Db 2 A 2

RESULT 5
BPP7_BOTIN STANDARD; PRT; 5 AA.
ID AC P30425;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Bradykinin-potentiating peptide S5,2 (5A) (Angiotensin-converting
DE enzyme inhibitor).
OS Bothrops insularis (Island jararaca) (Quelmada jararaca).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;  
 OC Viperidae; Crotalinae; Bothrops.  
 OX NCBI\_TaxID=8723;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Venom;  
 RA MEDLINE=90351557; PubMed=2386615;  
 RX Cintra A.C.O., Vieira C.A., Giglio J.R.;  
 RT "Primary structure and biological activity of bradykinin potentiating  
 peptides from Bothrops insularis snake venom.";  
 RL J. Protein Chem. 9:221-227(1990).  
 CC -|- FUNCTION: THIS PEPTIDE BOTH INHIBITS THE ACTIVITY OF THE  
 CC ANGIOTENSIN-CONVERTING ENZYME AND ENHANCES THE ACTION OF  
 CC BRADYKININ BY INHIBITING THE KINASES THAT INACTIVATE IT.  
 CC IT ACTS AS AN INDIRECT HYPOTENSIVE AGENT.  
 DR PIR: G37196; G37196  
 CC PIR: G37196; G37196  
 KW Hypotensive agent; Venom.  
 FT MODRES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 SEQUENCE 5 AA; 629 MW; 776DC37326B00000 CRC64;  
 Query Match 9.5%; Score 4; DB 1; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 18 A 18  
 Db 4 A 4  
 RESULT 6  
 RE11\_LITRU  
 ID RE11\_LITRU STANDARD; PRT; 5 AA.  
 AC P82070;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Rubellidin 1.1.  
 OS Litoria rubella (Desert tree frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Hylidae;  
 OC Pelodyadinae; Litoria.  
 OX NCBI\_TaxID=104895;  
 RN [1]  
 RP SEQUENCE, AND MASS SPECTROMETRY.  
 RC TISSUE=Skin secretion;  
 RA Steinborner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,  
 TYLER M.J., Wallace J.C.;  
 RT "The structure of new peptides from the Australian red tree frog  
 'Litoria rubella', the skin peptide profile as a probe for the study  
 of evolutionary trends of amphibians.";  
 RL Aust. J. Chem. 49:955-963(1996).  
 CC -|- FUNCTION: SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR ANTIBIOTIC  
 CC ACTIVITY.  
 CC -|- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
 CC -|- MASS SPECTROMETRY: MW=598; METHOD=FAB.  
 KW Amphibian skin.  
 SEQUENCE 5 AA; 598 MW; 6DD9C9CAB2A00000 CRC64;  
 Query Match 9.5%; Score 4; DB 1; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 18 A 18  
 Db 5 A 5  
 RESULT 7  
 RE21\_LITRU  
 ID RE21\_LITRU STANDARD; PRT; 5 AA.  
 AC P82071;  
 DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Rubellidin 2.1.  
 OS Litoria rubella (Desert tree frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Hylidae;  
 OC Pelodyadinae; Litoria.  
 OX NCBI\_TaxID=104895;  
 RN [1]  
 RP SEQUENCE, AND MASS SPECTROMETRY.  
 RC TISSUE=Skin secretion;  
 RA Steinborner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,  
 TYLER M.J., Wallace J.C.;  
 RT "The structure of new peptides from the Australian red tree frog  
 'Litoria rubella', the skin peptide profile as a probe for the study  
 of evolutionary trends of amphibians.";  
 RL Aust. J. Chem. 49:955-963(1996).  
 CC -|- FUNCTION: SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR ANTIBIOTIC  
 CC ACTIVITY.  
 CC -|- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
 CC -|- MASS SPECTROMETRY: MW=626; METHOD=FAB.  
 KW Amphibian skin.  
 SEQUENCE 5 AA; 626 MW; 6DD9C9CBI03000000 CRC64;  
 Query Match 9.5%; Score 4; DB 1; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 18 A 18  
 Db 5 A 5  
 RESULT 8  
 SUGA\_ACHDO  
 ID SUGA\_ACHDO STANDARD; PRT; 5 AA.  
 AC P19391;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Suboesophageal ganglion pentapeptide.  
 OS Acheta domesticus (House cricket).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Orthopteroidea; Orthoptera; Ensifera;  
 OC Gryllidae; Gryllinae; Acheta.  
 OX NCBI\_TaxID=6997;  
 RN [1]  
 RP SEQUENCE.  
 RA Wicker C., Wicker C.;  
 RT "Isolation and structure of a peptide isolated from the  
 suboesophageal ganglion of Acheta domesticus (orthoptera).";  
 RL Comp. Biochem. Physiol. 88C:185-187(1987).  
 CC -|- SUBCELLULAR LOCATION: MAIN PEPTIDE FROM THE SUBOESOPHAGEAL  
 CC GANGLIA.  
 DR PIR: JS0319; JS0319  
 SEQUENCE 5 AA; 476 MW; 69D76DDDDDB00000 CRC64;  
 Query Match 9.5%; Score 4; DB 1; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 18 A 18  
 Db 1 A 1  
 RESULT 9  
 CIP2\_MYED  
 ID CIP2\_MYED STANDARD; PRT; 6 AA.  
 AC P13737;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Contraction-inhibiting peptide II (MIP II).  
OS *Mytilus edulis* (Blue mussel).  
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;  
OC Mytiloidea; Mytilidae; *Mytilus*.  
OX NCBI\_TaxID=6550;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Fed. ganglion;  
RX MEDLINE=88240357; PubMed=3377776;  
RA Hirata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.;  
RT "Structures and actions of *Mytilus* inhibitory peptides.";  
RL Biochem. Biophys. Res. Commun. 152:1376-1382(1988).  
CC -1- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN  
CC MUSCLES.  
CC -1- SIMILARITY: TO MIP I.  
CC PIR; B27696; B27696.  
DR Hormone; Amidation.  
KW MOD\_RES 6  
FT MOD\_RES 6  
SQ SEQUENCE 6 AA; 621 MW; 72C9C6876DD81000 CRC64;  
  
Query Match 9.5%; Score 4; DB 1; Length 6;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 18 A 18  
Db 2 A 2  
  
RESULT 10  
LOK1.LOCMI STANDARD; PRT; 6 AA.  
ID LOK1.LOCMI  
AC P41491;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Locustakinin I.  
OS *Locusta migratoria* (Migratory locust).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;  
OC Acridomorpha; Acridoidea; Acrididae; Oedipodinae; Locustae.  
OX NCBI\_TaxID=7004;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Corpora cardiaca;  
RX MEDLINE=92262851; PubMed=1585017;  
RA Schoofs L., Holman G.M.; Proost P., van Damme J., Hayes T.K.,  
RA de Loof A.;  
RT "Locustakinin, a novel myotropic peptide from *Locusta migratoria*,  
RT isolation, primary structure and synthesis.";  
RL Regul. Pept. 37:49-57(1992).  
CC -1- FUNCTION: MYOTROPIC PEPTIDE. MAY BE IMPORTANT IN THE STIMULATION  
CC OF ION TRANSPORT AND INHIBITION OF DIURETIC ACTIVITY IN MALPIGHIAN  
CC TUBULES.  
CC PIR; A61068; A61068.  
DR Neuropeptide; Amidation.  
KW MOD\_RES 6  
FT MOD\_RES 6  
SQ SEQUENCE 6 AA; 654 MW; 686365A5B9CDB000 CRC64;  
  
Query Match 9.5%; Score 4; DB 1; Length 6;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 18 A 18  
Db 1 A 1  
  
RESULT 11  
OWN\_LEPDE STANDARD; PRT; 6 AA.  
ID OWN\_LEPDE  
AC P42985;  
DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Oviductal motility stimulating peptide (LeD-OVM).  
OS *Leptinotarsa decemlineata* (Colorado potato beetle).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;  
OC Cucujiformia; Phytophaga; Chrysomeloidea; Chrysomelidae;  
OC Chrysomelinae; *Leptinotarsa*.  
OX NCBI\_TaxID=7539;  
RN [1]  
RP SEQUENCE, AND SYNTHESIS.  
RC TISSUE=Head;  
RX MEDLINE=91271080; PubMed=2052497;  
RA Spittaels K., Schoofs L., Grauwels L., Smet H., van Damme J.,  
RA Proost P., Torrekens S., de Loof A.;  
RT "Isolation, identification and synthesis of novel oviductal motility  
RT stimulating head peptide in the Colorado potato beetle, *Leptinotarsa*  
RT *decemlineata*.";  
RL Peptides 12:31-36(1991).  
CC -1- FUNCTION: MYOTROPIC PEPTIDE. STIMULATES THE CONTRACTIONS OF THE  
CC OVIDUCT.  
CC Neuropeptide; Amidation.  
KW MOD\_RES 6  
FT MOD\_RES 6  
SQ SEQUENCE 6 AA; 720 MW; 6B07632B5DD03000 CRC64;  
  
Query Match 9.5%; Score 4; DB 1; Length 6;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 18 A 18  
Db 2 A 2  
  
RESULT 12  
TRPI\_PSEPU STANDARD; PRT; 6 AA.  
ID TRPI\_PSEPU  
AC P36414; 1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE TrpBA operon transcriptional activator (Fragment).  
GN TRPI.  
OS *Pseudomonas putida*.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC *Pseudomonas*.  
OX NCBI\_TaxID=303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PG1.CJS;  
RX MEDLINE=89335826; PubMed=2503057;  
RA Eberly L., Crawford I.P.;  
RT "DNA sequence of the tryptophan synthase genes of *Pseudomonas*  
RT *putida*.";  
RL Biochimie 71:521-531(1989).  
CC -1- FUNCTION: ACTIVATES THE EXPRESSION OF THE TRPBA GENES ENCODING  
CC THE TWO TRYPTOPHAN SYNTHASE SUBUNITS. IN THE ABSENCE OF THE  
CC INDUCER (INDOLEGLYCEROL PHOSPHATE), TRPI BINDS UPSTREAM OF THE  
CC TRPAB OPERON, OVERLAPPING ITS OWN PROMOTER REGION.  
CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X13299; CAA31660.1; -;  
DR InterPro; IPR000847; HTH\_LYSR.  
DR PROSITE; PS00044; HTH\_LYSR\_FAMILY; PARTIAL.







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 8, 2003, 10:54:50 ; Search time 23.1983 Seconds  
(without alignments)  
346.398 Million cell updates/sec

Title: US-09-003-869-5  
Perfect score: 42  
Sequence: 1 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 39

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mic.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID     | Description        |
|------------|-------|-------------|--------|-----------|--------------------|
| 1          | 4     | 9.5         | 5      | 10 Q99007 | Q99007 hordeum vul |
| 2          | 4     | 9.5         | 6      | 10 P82181 | P82181 spinacia ol |
| 3          | 4     | 9.5         | 6      | 10 P82182 | P82182 spinacia ol |
| 4          | 4     | 9.5         | 7      | 2 O50556  | O50556 actinobacil |
| 5          | 4     | 9.5         | 7      | 2 Q47477  | Q47477 escherichia |
| 6          | 4     | 9.5         | 7      | 2 Q47505  | Q47505 escherichia |
| 7          | 4     | 9.5         | 7      | 2 P72081  | P72081 nocardia la |
| 8          | 4     | 9.5         | 7      | 4 Q15897  | Q15897 homo sapien |
| 9          | 4     | 9.5         | 7      | 4 Q15903  | Q15903 homo sapien |
| 10         | 4     | 9.5         | 7      | 8 Q95945  | Q95945 saccharomyc |
| 11         | 4     | 9.5         | 7      | 11 O55184 | O55184 rattus norv |
| 12         | 4     | 9.5         | 7      | 11 Q63668 | Q63668 rattus norv |
| 13         | 4     | 9.5         | 7      | 12 Q66113 | Q66113 cherry leaf |
| 14         | 4     | 9.5         | 7      | 12 Q65578 | Q65578 bovine herp |
| 15         | 4     | 9.5         | 7      | 15 Q07624 | Q07624 rous sarcom |
| 16         | 4     | 9.5         | 8      | 2 O09258  | O09258 synchococc  |

|    |   |     |   |          |                    |
|----|---|-----|---|----------|--------------------|
| 17 | 4 | 9.5 | 8 | 2 O52062 | O52062 bacillus me |
| 18 | 4 | 9.5 | 8 | 2 O68485 | O68485 klebsiella  |
| 19 | 4 | 9.5 | 8 | 2 O9S6D5 | O9S6D5 escherichia |
| 20 | 4 | 9.5 | 8 | 2 O9S443 | O9S443 pseudomonas |
| 21 | 4 | 9.5 | 8 | 2 O32560 | O32560 escherichia |
| 22 | 4 | 9.5 | 8 | 2 P72221 | P72221 pseudomonas |
| 23 | 4 | 9.5 | 8 | 2 Q9R772 | Q9R772 escherichia |
| 24 | 4 | 9.5 | 8 | 2 Q49534 | Q49534 mycoplasma  |
| 25 | 4 | 9.5 | 8 | 2 Q56759 | Q56759 xanthobacte |
| 26 | 4 | 9.5 | 8 | 2 Q934S4 | Q934S4 thiobacillu |
| 27 | 4 | 9.5 | 8 | 2 P83152 | P83152 anabaena sp |
| 28 | 4 | 9.5 | 8 | 2 Q9R5R2 | Q9R5R2 shigella dy |
| 29 | 4 | 9.5 | 8 | 2 Q9R5R0 | Q9R5R0 shigella dy |
| 30 | 4 | 9.5 | 8 | 2 Q9R4M3 | Q9R4M3 enterococcu |
| 31 | 4 | 9.5 | 8 | 2 Q9R5L7 | Q9R5L7 clostridium |
| 32 | 4 | 9.5 | 8 | 2 Q33790 | Q33790 streptococc |
| 33 | 4 | 9.5 | 8 | 2 Q8RSR3 | Q8RSR3 lactobacill |
| 34 | 4 | 9.5 | 8 | 3 Q9HDS4 | Q9HDS4 aspergillus |
| 35 | 4 | 9.5 | 8 | 3 Q9URB9 | Q9URB9 saccharomyc |
| 36 | 4 | 9.5 | 8 | 3 P82858 | P82858 puccinia re |
| 37 | 4 | 9.5 | 8 | 4 Q9P285 | Q9P285 homo sapien |
| 38 | 4 | 9.5 | 8 | 4 Q9UH11 | Q9UH11 homo sapien |
| 39 | 4 | 9.5 | 8 | 4 Q9Y4X6 | Q9Y4X6 homo sapien |
| 40 | 4 | 9.5 | 8 | 4 Q15900 | Q15900 homo sapien |
| 41 | 4 | 9.5 | 8 | 4 Q15902 | Q15902 homo sapien |
| 42 | 4 | 9.5 | 8 | 4 Q96QF9 | Q96QF9 homo sapien |
| 43 | 4 | 9.5 | 8 | 4 Q9UCN4 | Q9UCN4 homo sapien |
| 44 | 4 | 9.5 | 8 | 4 Q60773 | Q60773 homo sapien |
| 45 | 4 | 9.5 | 8 | 5 Q9UB13 | Q9UB13 albinaria h |

## ALIGNMENTS

RESULT 1  
Q99007 ID Q99007 PRELIMINARY: PRT; 5 AA.  
AC Q99007;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-NOV-1996 (TREMBLrel. 08, Last annotation update)  
DE Alpha-amylase (EC 3.2.1.1) (Fragment).  
GN AMY1  
OS Hordeum vulgare (Barley).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
OC Triticeae; Hordeum.  
OX NCBI\_TaxID=4513;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HIMALAYA; TISSUE=ALEURONE LAYER;  
RX MEDLINE=91329704; PubMed=1831055;  
RA Jacobsen J.V., Close T.J.;  
RT "Control of transient expression of chimaeric genes by gibberellic  
RT acid and abscisic acid in protoplasts prepared from mature barley  
RT aleurone layers.";  
RL Plant Mol. Biol. 16:713-721(1991).  
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSTS OF 1,4-ALPHA-GLUCOSIDIC  
CC LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.  
CC -!- CORFACTOR: BINDS A CALCIUM ION REQUIRED FOR ITS ACTIVITY.  
CC -!- MISCELLANEOUS: THERE ARE AT LEAST 4 TYPES OF ALPHA-AMYLASE IN  
CC BARLEY.  
DR EMBL; X54643; CAA38455.1; -;  
KW Hydrolase; Glycosidase; Carbohydrate metabolism; Seed; Germination;  
FT Calcium; Multigene family.  
FT NON\_TER 5  
SQ SEQUENCE 5 AA; 600 MW; 61E3344DD6F00000 CRC64;

Query Match 9.5%; Score 4; DB 10; Length 5;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18

```
DB 2 A 2
RESULT 2
P82181 PRELIMINARY; PRT; 6 AA.
AC P82181: 2000 (TrEMBLrel. 14, Created)
DT 01-JUN-2000 (TrEMBLrel. 14, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Chloroplast 50S ribosomal protein L10 beta (Fragment).
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
OX NCBI_TaxID=3562;
RN [1]
RP SEQUENCE.
STRAIN=CV. ALVARO; TISSUE=LEAF;
MEDLINE=20435798; PubMed=10874046;
Yamaguchi K., Subramanian A.R.;
"The plastid ribosomal proteins. Identification of all the proteins in
the 50 S subunit of an organelle ribosome (chloroplast).";
J. Biol. Chem. 275:28466-28482(2000).
CC -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA.
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL PLANT TISSUES.
CC -!- MISCELLANEOUS: ON THE 2D-GEL ITS MW IS: 16.5 KDA.
CC -!- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
DR InterPro: IPR001790; Ribosomal_L10.
DR Pfam: PF00466; Ribosomal_L10; PARTIAL.
DR PROSITE: PS01109; RIBOSOMAL_L10; PARTIAL.
KW Ribosomal protein; Chloroplast; rRNA-binding.
FT NON_TER 6
SQ SEQUENCE 6 AA; 675 MW; 6321B415B05DB000 CRC64;

Query Match 9.5%; Score 4; DB 10; Length 6;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
DB 1 A 1

RESULT 3
P82182 PRELIMINARY; PRT; 6 AA.
AC P82182: 2000 (TrEMBLrel. 14, Created)
DT 01-JUN-2000 (TrEMBLrel. 14, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Chloroplast 50S ribosomal protein L10 gamma (Fragment).
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
OX NCBI_TaxID=3562;
RN [1]
RP SEQUENCE.
STRAIN=CV. ALVARO; TISSUE=LEAF;
MEDLINE=20435798; PubMed=10874046;
Yamaguchi K., Subramanian A.R.;
"The plastid ribosomal proteins. Identification of all the proteins in
the 50 S subunit of an organelle ribosome (chloroplast).";
J. Biol. Chem. 275:28466-28482(2000).
CC -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA.
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL PLANT TISSUES.
CC -!- MISCELLANEOUS: ON THE 2D-GEL ITS MW IS: 16.5 KDA.
CC -!- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
DR InterPro: IPR001790; Ribosomal_L10.
DR Pfam: PF00466; Ribosomal_L10; PARTIAL.
DR PROSITE: PS01109; RIBOSOMAL_L10; PARTIAL.
KW Ribosomal protein; Chloroplast; rRNA-binding.
FT NON_TER 6
SQ SEQUENCE 6 AA; 675 MW; 6321B415B05DB000 CRC64;

Query Match 9.5%; Score 4; DB 10; Length 6;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
DB 1 A 1

RESULT 3
P82182 PRELIMINARY; PRT; 6 AA.
AC P82182: 2000 (TrEMBLrel. 14, Created)
DT 01-JUN-2000 (TrEMBLrel. 14, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Chloroplast 50S ribosomal protein L10 gamma (Fragment).
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
OX NCBI_TaxID=3562;
RN [1]
RP SEQUENCE.
STRAIN=CV. ALVARO; TISSUE=LEAF;
MEDLINE=20435798; PubMed=10874046;
Yamaguchi K., Subramanian A.R.;
"The plastid ribosomal proteins. Identification of all the proteins in
the 50 S subunit of an organelle ribosome (chloroplast).";
J. Biol. Chem. 275:28466-28482(2000).
CC -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA.
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL PLANT TISSUES.
CC -!- MISCELLANEOUS: ON THE 2D-GEL ITS MW IS: 16.5 KDA.
CC -!- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
DR InterPro: IPR001790; Ribosomal_L10.
DR Pfam: PF00466; Ribosomal_L10; PARTIAL.
DR PROSITE: PS01109; RIBOSOMAL_L10; PARTIAL.
KW Ribosomal protein; Chloroplast; rRNA-binding.
FT NON_TER 6
SQ SEQUENCE 6 AA; 675 MW; 6321B415B05DB000 CRC64;
```

```
DR InterPro: IPR001790; Ribosomal_L10.
DR InterPro: IPR002363; Ribosomal_L10eub.
DR Pfam: PF00466; Ribosomal_L10; PARTIAL.
DR PROSITE: PS01109; RIBOSOMAL_L10; PARTIAL.
KW Ribosomal protein; Chloroplast; rRNA-binding.
FT NON_TER 6
SQ SEQUENCE 6 AA; 675 MW; 6321B415B05DB000 CRC64;

Query Match 9.5%; Score 4; DB 10; Length 6;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
DB 1 A 1

RESULT 4
O50556 PRELIMINARY; PRT; 7 AA.
AC O50556: 1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GiYA (Fragment).
GN GiYA.
OS Actinobacillus actinomycetemcomitans (Haemophilus
OC actinomycetemcomitans).
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Actinobacillus.
OX NCBI_TaxID=714;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33584;
RX MEDLINE=96355846; PubMed=8751884;
RA Kolodrubetz D., Spitznagel J. Jr., Wang B., Phillips L.H., Jacobs C.,
RA Kraig E.;
RT "cis Elements and trans factors are both important in strain-specific
RT regulation of the leukotoxin gene in Actinobacillus
RL Infect. Immun. 64:3451-3460(1996).
DR EMBL: U51862; AAB88721.1; -.
FT NON_TER 1
SQ SEQUENCE 7 AA; 832 MW; 6DCB42D767340420 CRC64;

Query Match 9.5%; Score 4; DB 2; Length 7;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
DB 7 A 7

RESULT 5
O47477 PRELIMINARY; PRT; 7 AA.
AC O47477: 1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE Tpi protein (Fragment).
GN TPI.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE OF 7-7 FROM N.A.
RX MEDLINE=85203917; PubMed=3158524;
RA Hellinga H.W., Evans P.R.;
RT "Nucleotide sequence and high-level expression of the major
RT Escherichia coli phosphofructokinase.";
```

```
RL Eur. J. Biochem. 149:363-373(1985).
RN [2]
RP SEQUENCE FROM N.A.
RA Evans P.;
RL Submitted (OCT-1986) to the EMBL/GenBank/DBJ databases.
DR EMBL: X02519; CAA26359.1; -.
FT NON_TER 1
SQ SEQUENCE 7 AA; 773 MW; 7416D33DDDB1DB0 CRC64;

Query Match
Best Local Similarity 9.5%; Score 4; DB 2; Length 7;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
DB 1 A 1

RESULT 6
Q7505 PRELIMINARY; PRT; 7 AA.
Q47505;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE MCCA protein.
GN MCCA.
OS Escherichia coli.
OG Plasmid pMccC7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96099297; PubMed=8522520;
RA Gonzalez-Pastor J.E., San Millan J.L., Castilla M.A., Moreno F.;
RT "Structure and organization of plasmid genes required to produce the
RT translation inhibitor microcin C7.";
RL J. Bacteriol. 177:7131-7140(1995).
DR EMBL: X57583; CAA40808.1; -.
KW Plasmid.
SQ SEQUENCE 7 AA; 763 MW; 644DD4861B406F0 CRC64;

Query Match
Best Local Similarity 9.5%; Score 4; DB 2; Length 7;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

18 A 18
6 A 6

RESULT 7
P72081 PRELIMINARY; PRT; 7 AA.
ID AC P72081;
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE 3'-methylcephem hydroxylase (Fragment).
GN CEFF.
OS Nocardia lactamdurans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.
OX NCBI_TaxID=1913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96009872; PubMed=7557411;
RA Coque J., Perez-Llarena F.J., Enguita F.J., Fuente J.L., Martin J.F.,
RA Miras P.;
RT "Characterization of the cmcH genes of Nocardia lactamdurans and
RT Streptomyces clavuligerus encoding a functional 3'-hydroxymethylcephem
RT O-carbamoyltransferase for cephamycin biosynthesis.";
```

```
RL Gene 162:21-27(1995).
DR EMBL: Z21682; CAA79797.1; -.
FT NON_TER 1
SQ SEQUENCE 7 AA; 746 MW; 75A1B2CDD1E681C0 CRC64;

Query Match
Best Local Similarity 9.5%; Score 4; DB 2; Length 7;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 A 18
DB 4 A 4

RESULT 8
Q15897 PRELIMINARY; PRT; 7 AA.
AC Q15897;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE (clone XP6A11A) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Lee C.-C., Yazdani A., Wehnert M., Bailey J., Couch L., Xiong M.,
RA Coolbaugh M.I., Chinault C.A., Baldini A., Lindsay E.A., Zhao Z.-Y.,
RA Caskey C.T.H.;
RT "Isolation of chromosome-specific genes by reciprocal probing of
RT arrayed cDNAs and cosmid libraries.";
RL Hum. Mol. Genet. 0:0-0(1995).
DR EMBL: L32077; AAA73887.1; -.
FT NON_TER 1
FT NON_TER 7
SQ SEQUENCE 7 AA; 814 MW; 672B1DD3372046B0 CRC64;

Query Match
Best Local Similarity 9.5%; Score 4; DB 4; Length 7;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

18 A 18
5 A 5

RESULT 9
Q15903 PRELIMINARY; PRT; 7 AA.
AC Q15903;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE (clone XP7E7B) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Lee C.-C., Yazdani A., Wehnert M., Bailey J., Couch L., Xiong M.,
RA Coolbaugh M.I., Chinault C.A., Baldini A., Lindsay E.A., Zhao Z.-Y.,
RA Caskey C.T.H.;
RT "Isolation of chromosome-specific genes by reciprocal probing of
RT arrayed cDNAs and cosmid libraries.";
RL Hum. Mol. Genet. 0:0-0(1995).
DR EMBL: L32082; AAA73893.1; -.
FT NON_TER 1
FT NON_TER 7
```

SQ SEQUENCE 7 AA; 849 MW; 6B04039CDD33DB0 CRC64;  
Query Match 9.5%; Score 4; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 18 A 18  
Db 1 A 1  
RESULT 10  
Q95945 PRELIMINARY; PRT; 7 AA.  
AC Q95945;  
DT 01-FEB-1997 (TReMBLrel. 02, Created)  
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)  
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Inside intron 5 (fragment).  
Saccharomyces cerevisiae (Baker's yeast).  
Mitochondrion.  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D273-10B;  
RX MEDLINE=81069885; PubMed=6254986;  
RA Bonitz S.G., Coruzzi G., Thalenfeld B., Tzagoloff A., Macino G.;  
RT "Assembly of the mitochondrial membrane system: Structure and  
RT nucleotide sequence of the gene coding for subunit 1 of yeast  
RT cytochrome oxidase".  
RL J. Biol. Chem. 255:11927-11941(1980).  
DR EMBL; V00694; CAA24066.1; -.  
KW Mitochondrion.  
FT NON\_TER 1  
SQ SEQUENCE 7 AA; 859 MW; 75B7232362CDC460 CRC64;  
Query Match 9.5%; Score 4; DB 8; Length 7;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 18 A 18  
Db 2 A 2  
RESULT 11  
Q55184 PRELIMINARY; PRT; 7 AA.  
AC Q55184;  
DT 01-JUN-1998 (TReMBLrel. 06, Created)  
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE Orphan receptor TR4-NS (Fragment).  
GN TR4.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY;  
RX MEDLINE=96198747; PubMed=8612486;  
RA Yoshikawa T., Makino S., Gao X.M., Xing G.O., Chuang D.M.,  
RA Detera-Wadleigh S.D.;  
RT "Splice variants of rat TR4 orphan receptor: differential expression  
RT of novel sequences in the 5'-untranslated region and C-terminal  
RT domain".  
RL Endocrinology 137:1562-1571(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96124520; PubMed=8560786;  
RA Borja M., Sanchez F., Rowhani A., Bruening G., Ponz F.;

RX MEDLINE=96299786; PubMed=8661150;  
RA Yoshikawa T., DuPont B.R., Leach R.J., Detera-Wadleigh S.D.;  
RT "New variants of the human and rat nuclear hormone receptor, TR4:  
RT expression and chromosomal localization of the human gene".  
RL Genomics 35:361-366(1996).  
DR EMBL; U59454; AAB91433.1; -.  
KW Receptor.  
FT NON\_TER 1  
SQ SEQUENCE 7 AA; 663 MW; 6DDAA8787EB05350 CRC64;  
Query Match 9.5%; Score 4; DB 11; Length 7;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 18 A 18  
Db 7 A 7  
RESULT 12  
Q63668 PRELIMINARY; PRT; 7 AA.  
AC Q63668;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE ORF2 protein.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=KIDNEY;  
RX MEDLINE=95396550; PubMed=7667072;  
RA Mandon B., Bellanger A.C., Elalouf J.M.;  
RT "Inverse-PCR-mediated cloning of the promoter for the rat vasopressin  
RT V2 receptor gene".  
RL Pflugers Arch. 430:12-18(1995).  
DR EMBL; X83264; CAA58237.1; -.  
SQ SEQUENCE 7 AA; 703 MW; 75A767287DC6D6F0 CRC64;  
Query Match 9.5%; Score 4; DB 11; Length 7;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 18 A 18  
Db 3 A 3  
RESULT 13  
Q66113 PRELIMINARY; PRT; 7 AA.  
AC Q66113;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE C-terminus of the viral replicase (Fragment).  
OS cherry leaf roll virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;  
OC Nepovirus.  
OX NCBI\_TaxID=12615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WALNUT;  
RA Borja M.;  
RL Thesis (1992), Biologia Molecular y Virologia Vegetal, CIT-INIA.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WALNUT;  
RX MEDLINE=96124520; PubMed=8560786;  
RA Borja M., Sanchez F., Rowhani A., Bruening G., Ponz F.;

RT "Long, nearly identical untranslated sequences at the 3' terminal  
 RT regions of the genomic RNAs of cherry leafroll virus (walnut  
 strain).";  
 RL Virus Genes 10:245-252(1995).  
 DR EMBL: Z34265; CAA84019.1; -.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 7 AA: 800 MW; 7417672BDC6D740 CRC64;

Query Match 9.5%; Score 4; DB 12; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 18 A 18  
 |  
 Db 3 A 3

## RESULT 14

Q65578 PRELIMINARY; PRT; 7 AA.

Q65578;  
 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Hypothetical 0.8 kDa protein (Fragment).  
 OS Bovine herpesvirus 1.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Varicellovirus.  
 OX NCBI\_TaxID=10320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COOPER;  
 RA MEDLINE=95313343; PubMed=7793062;  
 RA Vleck C., Benes V., Lu Z., Kutish G.F., Paces V., Rock D.,  
 RA Letchworth G.J., Schwzyer M.;  
 RT "Nucleotide sequence analysis of a 30-kb region of the bovine  
 RT herpesvirus 1 genome which exhibits a colinear gene arrangement with  
 RT the UL21 to UL4 genes of herpes simplex virus.";  
 RL Virology 210:100-108(1995).  
 DR EMBL: Z48053; CAA88130.1; -.  
 KW Hypothetical protein.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 7 AA: 758 MW; 6DD33455B1F1B1C0 CRC64;

Query Match 9.5%; Score 4; DB 12; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

18 A 18  
 |  
 Db 7 A 7

## RESULT 15

Q07624 PRELIMINARY; PRT; 7 AA.

Q07624;  
 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Rous sarcoma virus.  
 OC Viruses; Retroviridae; Alpharetrovirus.  
 OX NCBI\_TaxID=11886;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PRAGUE C;  
 RX MEDLINE=93010967; PubMed=1327749;  
 RA Donze O., Spahr P.F.;  
 RT "Role of the open reading frames of Rous sarcoma virus leader RNA in  
 RT translation and genome packaging."  
 RL EMBO J. 11:3747-3757(1992).  
 DR EMBL: X67587; CAA47862.1; -.

KW Hypothetical protein.  
 FT NON\_TER 7 7  
 SQ SEQUENCE 7 AA: 672 MW; 776045A7687DD6F0 CRC64;

Query Match 9.5%; Score 4; DB 15; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 18 A 18  
 |  
 Db 2 A 2

Search completed: January 8, 2003, 11:00:53  
 Job time : 25.1983 secs

